



IDENTIFICACIÓN EN LÍNEA DE ÁREAS CON PROBLEMAS DE ESTABILIDAD DE TENSIÓN EN SISTEMAS ELÉCTRICOS DE POTENCIA MODERNOS

Juan David Angarita Colorado

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERIAS

PROGRAMA DE INGENIERÍA ELÉCTRICA

PEREIRA

2015

**IDENTIFICACIÓN EN LÍNEA DE ÁREAS CON PROBLEMAS DE ESTABILIDAD
DE TENSIÓN EN SISTEMAS ELÉCTRICOS DE POTENCIA MODERNOS**

Juan David Angarita Colorado

Proyecto de grado para optar al título de Ingeniero Eléctricista

Director: Juan José Mora Flórez, PhD.

UNIVERSIDAD TECNOLÓGICA DE PEREIRA

FACULTAD DE INGENIERIAS

PROGRAMA DE INGENIERÍA ELÉCTRICA

PEREIRA

2015

Agradecimientos

A mi mamá y mi abuela, por brindarme incondicionalmente su amor y apoyo; por acompañarme durante todos estos años y ayudarme a crecer como persona, siempre mostrándome el lado bonito y especial que tiene la vida. A mi padre, por su apoyo para culminar esta etapa de vida. A mi hermana, por demostrarme que las adversidades de la vida pueden ser superadas. A mis abuelos, por su acompañamiento y cariño. A mi familia.

Al ingeniero Juan José Mora, por brindarme la oportunidad de trabajar con él, y por todo ese conocimiento y experiencia compartida durante este recorrido.

A Edwin Mauricio Ríos por su compañía infalible, por su apoyo incondicional y su cariño constante.

Al ingeniero Luis Fernando Rodríguez, por su disposición para compartir su conocimiento, sus recomendaciones y su asesoría para poder desarrollar con éxito este trabajo y más valioso aun por su amistad.

A mis compañeros, que estuvieron durante todo el proceso de formación y que de cierta forma me acompañaron para llegar hasta este punto de mi vida.

A la ingeniera Sandra Pérez y al ingeniero Andrés Zapata por las asesorías y explicaciones.

A las demás personas que directa o indirectamente han pasado por mi vida.

A todos, les quedo inmensamente agradecido.

Resumen

Para mantener el balance entre la generación y la demanda en los sistemas eléctricos de potencia modernos, y así mantener su estabilidad, se requiere la implementación de nuevas estrategias de monitoreo e identificación de áreas críticas, para así determinar las mejores acciones de control preventivo y/o correctivo necesarias.

Debido a lo anteriormente expuesto, en este trabajo de grado se propone una metodología de identificación de áreas críticas en relación a su estabilidad de tensión. Se propone la utilización de técnicas de agrupamiento, para determinar las áreas críticas que describen el sistema y el uso de una técnica de clasificación para la identificación en línea de condiciones de operación problemáticas.

La evaluación de la efectividad de la metodología propuesta, se realiza en un sistema de prueba tipo IEEE de 14 barras. A partir de los resultados obtenidos se espera éstos sean igualmente satisfactorios en otros sistemas de potencia.

Palabras clave: Aprendizaje de máquina, estabilidad de tensión, técnica de agrupación *k means*, técnica de clasificación *knn*.

Abstract

To maintain the balance between generation and load in the modern electric power system, and, as a direct consequence, assure its stability, the implementation of new strategies for monitoring and identification of critical areas are required, in order to perform the necessary preventive or corrective control actions.

Due to the above mentioned, it is proposed a methodology for identification of critical areas concerning their voltage stability. A clustering technique is proposed to determine the critical areas which describe the power system. A classification technique for online identification of problematic operating conditions is also proposed.

The proposed methodology is evaluated on an IEEE 14 bus test system. From the obtained results, it is expected an adequate performance in other power systems.

Keywords: Learning machine, voltage stability, the *k-means* clustering technique, the *knn* classification technique.

Contenido

Resumen.....	IV
Abstract	V
Lista de Figuras	VIII
Lista de Tablas	IX
1. Introducción.....	1
1.1 Definición del problema.....	1
1.2 Objetivos de la tesis.....	2
1.3 Estructura del documento	3
2. Aspectos teóricos.....	4
2.1 Estabilidad de tensión.....	4
2.1.1 Análisis de estabilidad de tensión	4
2.1.2 Índices de estabilidad de tensión.....	4
2.2 Minería de datos	7
2.2.1 Algoritmo de agrupación <i>k means</i>	7
2.2.2 Algoritmo de clasificación <i>k</i> vecinos más cercanos.....	8
2.2.3 Formas para calcular la distancia	10
3. Metodología para determinación de áreas críticas de estabilidad de tensión utilizando aprendizaje de máquina.....	11
3.1 Introducción.....	11
3.2 Modelado y simulación del sistema eléctrico de potencia	11
3.3 Generación base de datos	12
3.3.1 Latin Hypercube.....	12
3.3.2 Solución del flujo de carga.....	13
3.3.3 Índices de estabilidad de tensión.....	14
3.3.4 Descriptores utilizados y almacenamiento de la base de datos	15
3.4 Técnica de agrupación basada en el método <i>k means</i>	16
3.4.1 Estrategias para la evaluación del agrupamiento	20
3.4.2 Determinación de áreas críticas de estabilidad	24

3.5	Técnica de clasificación basada en el método k vecinos más cercanos	27
3.5.1	Entrenamiento del algoritmo k vecinos más cercanos	28
3.5.2	Parametrización del algoritmo k vecinos más cercanos y clasificación de una nueva condición de operación del sistema.....	29
4.	Aplicación de la metodología propuesta.....	31
4.1	Descripción del sistema de prueba IEEE de 14 barras	31
4.2	Descripción de las condiciones de operación para la validación de la propuesta .	32
4.2.1	Prueba en el sistema.....	32
4.2.2	Descriptores utilizados en la base de datos	32
4.3	Determinación de las áreas críticas de estabilidad	33
4.3.1	Matriz de condiciones de operación.....	33
4.3.2	Selección de los parámetros de la técnica de agrupamiento	33
4.3.3	Determinación del mejor agrupamiento.....	38
4.4	Reconocimiento de áreas críticas de estabilidad	38
4.4.1	Conjunto de entrada	38
4.4.2	Selección de los parámetros de la técnica de clasificación	39
4.4.3	Validación cruzada.....	39
4.4.4	Modificaciones al algoritmo básico	40
4.4.5	Parametrización del algoritmo k vecinos más cercanos.....	42
4.4.6	Evaluación del clasificador knn parametrizado.....	43
4.4.7	Comparación de la técnica de clasificación knn con la máquina de soporte vectorial	43
5.	Conclusiones y recomendaciones	45
5.1	Conclusiones.....	45
5.2	Recomendaciones	46
5.3	Trabajos futuros.....	47
	Bibliografía	48

Lista de Figuras

Figura 2.1: Representación simplificada de un sistema de potencia	5
Figura 2.2: Metodología de agrupación en dos dimensiones.....	8
Figura 2.3: Metodología de clasificación en dos dimensiones	9
Figura 3.1: Esquema de condiciones de operación para un sistema de potencia.....	14
Figura 3.2: Comportamiento del índice SVSI para la barra 12 considerando algunas condiciones de operación.	15
Figura 3.3: Esquema de agrupación mediante la técnica <i>k means</i>	17
Figura 3.4: Agrupamiento realizado mediante la técnica <i>k means</i>	18
Figura 3.5: Ejemplo de las condiciones de operación que conforman un grupo G_t y su prototipo representativo.	19
Figura 3.6: Condición de operación 1.....	21
Figura 3.7: Condición de operación 2.....	21
Figura 3.8: Diagrama de flujo para determinación de áreas críticas de estabilidad en tensión.	26
Figura 3.9: Esquema de clasificación mediante la técnica <i>knn</i>	27
Figura 3.10: Esquema de clasificación mediante la técnica <i>knn</i>	30
Figura 4.1: Diagrama unifilar del sistema IEEE de 14 barras	31
Figura 4.2: Índice Davis Bouldin para cada uno de los agrupamientos variando la forma de calcular la distancia	34
Figura 4.3: Comportamiento del índice Davis Bouldin para cada uno de los agrupamientos	35
Figura 4.4: Comportamiento del índice de diferencias de tensión para cada uno de los agrupamientos	36
Figura 4.5: Comportamiento cualitativo según las zonas de severidad definida para cada uno de los agrupamientos.....	37
Figura 4.6: Subconjuntos creados para la evaluación del clasificador mediante validación cruzada	40
Figura 4.7: Desempeño del clasificador	41
Figura 4.8: Desempeño del clasificador con los dos mejores parametros	42

Lista de Tablas

Tabla 3.1. Descriptores utilizados para describir cada condición de operación.	16
Tabla 3.2: Diferencias de tensión para los nt nodos que pertenecen a determinada zona de severidad Z_s para cada una de las condiciones de operación.....	23
Tabla 3.3: Comparación entre dos centroides que pertenecen a dos grupos diferentes; al lado izquierdo se observan los valores obtenidos de los índices, y al lado derecho se observa su equivalente cualitativo.	24
Tabla 4.1: Rangos de criticidad para definir las zonas de severidad	37
Tabla 4.2: Descriptores utilizados para describir las condiciones de operación en el proceso de clasificación.....	39
Tabla 4.3. Especificaciones del equipo de computo.	43

1. Introducción

En este capítulo se presenta la definición del problema que se trabaja en esta tesis, los objetivos que tiene el trabajo de investigación y finalmente, se presenta la estructura del documento.

1.1 Definición del problema

La estabilidad de tensión, es un punto clave en el análisis de los sistemas eléctricos de potencia, debido al gran esfuerzo que se hace en los centros de control, por mantener el balance entre la generación y la demanda. Actualmente, la operación del sistema de potencia se realiza en los centros de control, teniendo en cuenta los escenarios de operación planeados, la topología actual del sistema, ya que los elementos imponen los límites, y adicionalmente con los resultados obtenidos de ejecutar flujos de carga en software especiales de simulación (Pérez Londoño, 2013) (Congreso de la Republica de Colombia, 1994) (Congreso de la República de Colombia, 1994).

Debido a restricciones y limitaciones de la red eléctrica, se deben ejercer acciones de control ante los eventos o contingencias, tales como: variación en el nivel de la demanda, pérdida de unidades generadoras, salidas de líneas de transmisión, entre otras. Por lo tanto, se requiere información actualizada en tiempo real, para la adecuada operación del sistema eléctrico.

Como consecuencia de la inclusión de nuevas tecnologías al sistema, como las unidades de medición fasorial PMU (elementos de medición de tensión y corriente tanto en magnitud como en ángulo), el mejoramiento de los procesadores de los sistemas de cómputo, y la necesidad de mantener la estabilidad en el sistema, se requiere el desarrollo de herramientas para el monitoreo y análisis de la estabilidad del sistema de potencia en tiempo real (Pérez Londoño, 2013) (NYSERDA, 2010).

Sin embargo, debido a la gran complejidad del sistema de potencia en cuanto a su tamaño, abundantes interconexiones y la gran cantidad de elementos involucrados, el análisis de la estabilidad de tensión se vuelve una tarea compleja. Adicionalmente, la gran cantidad de condiciones de operación que se presentan, y como consecuencia, el manejo de todas las variables que se obtienen de los elementos de medición, incrementa el esfuerzo computacional. Un reto actual consiste en buscar nuevas herramientas, que logren identificar con alta precisión y en tiempo real, los estados de operación del sistema.

La identificación de áreas críticas, con relación a su estabilidad de tensión, es un avance en la forma cómo se opera el sistema desde el centro nacional. La identificación de las áreas que están presentando problemas, en relación con su estabilidad de tensión, beneficia la

operación, ya que agiliza las acciones de control a realizar en el sistema, ante una condición crítica de operación. Para crear protocolos de control, es necesario que el operador del sistema de potencia disponga de todas las herramientas adecuadas para tomar decisiones, ya que una acción incorrecta podría tener un impacto devastador, por ejemplo, dejar un país sin servicio de energía eléctrica.

Para identificar cuáles son las posibles condiciones del sistema de potencia en las cuales se deberían tomar diferentes acciones de control, se hace necesario adaptar algoritmos que presenten mejor desempeño y un costo computacional bajo, para así reducir al máximo el tiempo entre el inicio de la contingencia, la identificación, el procesamiento y la acción de control.

La operación del sistema se ha realizado a lo largo de los años de una forma adecuada y precisa, sin embargo, se han podido observar que diferentes *blackouts* han afectado a algunos países, incluido el nuestro; por lo tanto, no es posible asegurar que ante una futura contingencia, el sistema continúe estable (U S -Canada Power System Outage Task, 2011). Por lo planteado anteriormente, la evolución de los sistemas de monitoreo y de control, debe ser continua, para así garantizar que la confiabilidad del suministro no se vea afectada negativamente, debido al impacto que presenta el mercado de la energía eléctrica, en diversos sectores.

Esta investigación plantea la utilización del toolbox PSAT (Milano, 2001) para la simulación de escenarios de operación del sistema eléctrico de potencia. Utilizar métodos estadísticos reconocidos, como el muestreo Latin Hypercube, para generar las diferentes condiciones de operación del sistema. Mediante algoritmos de agrupación y de clasificación, de fácil implementación y entendimiento, los cuales han dado buenos resultados en otras áreas como la localización de fallas (Mora Floréz, 2006), se propone una metodología de identificación en línea de la condición actual del sistema. El desarrollo de herramientas veloces y con alta eficiencia, son esenciales en la solución de problemas de estabilidad dentro del sistema, ya que una acción de control que no se realice a tiempo, o se realice de forma inadecuada, puede ser la diferencia entre la pérdida de la estabilidad de tensión y/o la pérdida de elementos del sistema.

1.2 Objetivos de la tesis

Como objetivo general de la tesis se tiene, desarrollar una metodología de aplicación en línea para la determinación de áreas críticas, en términos de estabilidad de tensión, bajo diferentes condiciones de operación del sistema eléctrico de potencia.

Los objetivos específicos relacionados son:

- a. Revisar la bibliografía vigente para escoger las mejores herramientas desarrolladas y así adaptarlas a la solución del problema actual.

- b. Utilizar el toolbox PSAT de MATLAB para obtener las diferentes condiciones de operación, para así obtener diferentes condiciones que representen el sistema eléctrico de potencia.
- c. Implementar una estrategia de agrupación de áreas según su estabilidad de tensión.
- d. Implementar una estrategia de clasificación de áreas según su estabilidad de tensión.
- e. Evaluar el desempeño de la metodología implementada en un sistema de prueba tipo IEEE.
- f. Reportar los resultados de la investigación.

1.3 Estructura del documento

El documento presenta la siguiente estructura:

En el capítulo dos, se presentan los aspectos teóricos sobre estabilidad de tensión, y la forma de describir la estabilidad de tensión, de acuerdo a índices ampliamente utilizados; adicionalmente, se mencionan aspectos sobre la minería de datos; algoritmos de aprendizaje de máquina, para agrupación y clasificación de escenarios. Estos algoritmos son necesarios para implementar herramientas de monitoreo e identificación de condiciones de operación del sistema eléctrico de potencia.

Por otro lado, en el capítulo tres, se aborda una metodología de identificación en línea de áreas sensibles en relación a su estabilidad de tensión. Se utilizan los índices de tensión de línea y de barra, y mediante técnicas de agrupación y clasificación, se pretende reconocer en tiempo real, las áreas del sistema que requieren un control preventivo y/o correctivo para evitar un colapso en los niveles de tensión.

Posteriormente, en el capítulo cuatro, se aplica la metodología de identificación de áreas críticas en relación a la estabilidad de tensión sobre un sistema de prueba tipo IEEE, para así, analizar resultados y evaluar la efectividad de la metodología propuesta.

Para finalizar, se presentan las conclusiones más relevantes sobre la metodología implementada y los resultados observados en el sistema de prueba; también, se hacen recomendaciones sobre los algoritmos implementados y la forma de parametrizarlos.

2. Aspectos teóricos

En este capítulo, se presentan conceptos básicos sobre estabilidad de tensión en sistemas eléctricos de potencia y sobre métodos de aprendizaje de máquina.

2.1 Estabilidad de tensión

La estabilidad de tensión, se relaciona con la capacidad que posee un sistema de potencia para mantener niveles de tensión aceptados, en los nodos que lo componen, durante su operación en estado permanente. Otra definición de la estabilidad de tensión, es la capacidad que posee un sistema de volver a un estado de operación, con niveles aceptados, después de ser sometido a una perturbación (Kundur, 1994).

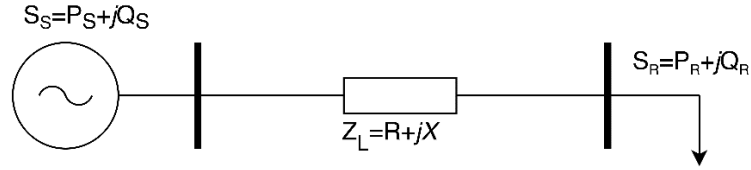
Se puede relacionar entonces la inestabilidad de tensión como un problema de orden local, ya que es la incapacidad del sistema de generar y/o transportar la energía reactiva que es demandada por el sistema; sin embargo, por la interconexión del sistema, este fenómeno puede traer consecuencias a niveles globales (Kundur, 1994). Debido a que la estabilidad de tensión puede ser vista como un fenómeno local, es importante reconocer cuales son las barras con mayor inestabilidad en el sistema para así realizar las acciones necesarias ante dicha condición de operación y evitar o corregir un colapso de tensión (Pérez Londoño, 2013).

2.1.1 Análisis de estabilidad de tensión

Debido al comportamiento del sistema, se pueden seguir metodologías de análisis estático o dinámico para determinar la cercanía del sistema a un estado de inestabilidad (Kundur, 1994). El análisis estático, se realiza para condiciones de carga específicas y se pondera por medio de factores de sensibilidad, calculados a partir de los resultados obtenidos al realizar un flujo de carga. El análisis dinámico, considera modelos más complejos de los elementos que componen el sistema, por lo cual, se hace más compleja su aplicación en sistemas de control en tiempo real, sin embargo, ambos se complementan para el análisis de la estabilidad de tensión del sistema. (Machowski, Bialek, & Bumby, 2008)

2.1.2 Índices de estabilidad de tensión

Los factores de sensibilidad de tensión, para identificar y clasificar los elementos que componen un sistema, con base en su estabilidad de tensión, como el mostrado en la Figura 2.1, son los índices de barra y de línea.

Figura 2.1: Representación simplificada de un sistema de potencia

2.1.2.1 Índices de estabilidad de barra

a. Índice L (Kessel & Glavitsch, 1986)

El índice L, utiliza la matriz de impedancia del sistema, reformulada como se muestra en la expresión (2.1), en donde se discriminan barras generadoras G y barras no generadoras L .

$$\begin{bmatrix} V_L \\ I_G \end{bmatrix} = \begin{bmatrix} Z_{LL} & F_{LG} \\ K_{GL} & Y_{GG} \end{bmatrix} \begin{bmatrix} I_L \\ V_G \end{bmatrix} \quad (2.1)$$

$$F_{LG} = -Y_{LL}^{-1} Y_{LG} \quad (2.2)$$

El índice L para una barra j , se define utilizando la expresión (2.3).

$$L_j = \left| 1 - \frac{\sum_{i=1}^{\#G} F_{ji} * V_i}{V_j} \right| \quad (2.3)$$

b. Índice VCPI (Balamourgan, Sidhu, & Sachdev, 2004)

El índice VCPI, utiliza la información de la tensión en forma fasorial de cada barra y la matriz de admitancia de la red. Este puede ser hallado por medio de la expresión (2.4), donde k representa cada barra del sistema.

$$VCPI_k = 1 - \left| \frac{\sum_{m=1, m \neq k}^N \frac{Y_{km}}{\sum_{j=1, j \neq k}^N Y_{kj}} V_m}{V_k} \right| \quad (2.4)$$

c. Índice SVSI (Pérez Londoño, Rodríguez García, & Olivar, 2014)

El índice SVSI, se basa en la distancia eléctrica relativa entre las barras generadoras y de carga, y contempla factores de sensibilidad utilizando las variables del sistema eléctrico de potencia, se plantea de la siguiente manera:

Utilizando las expresiones (2.1) y (2.2), se calcula la distancia relativa R_{LG} , definida en (Yesuratnam & Thukaram, 2007) y se calcula mediante la expresión (2.5)

$$R_{LG} = [A] - abs(F_{LG}) \quad (2.5)$$

Donde A se define como una matriz de dimensiones $(n-g) \times g$, n es el número total de barras del sistema y g el número de generadores.

Después de encontrar la matriz de distancias eléctricas relativas, se selecciona la barra donde está el generador más cercano a la barra estudiada, y se realiza la siguiente aproximación para la caída de tensión.

$$\Delta V_i^j(t) = \sum_{b=1}^{n_j-1} |\vec{V}_b - \vec{V}_{b+1}| \cong |\vec{V}_g - \vec{V}_l| \quad (2.6)$$

Donde V_g y V_l corresponden a las tensiones de la barra donde se encuentra el generador más cercano y de la barra analizada, respectivamente. Para el cálculo del índice, se tiene en cuenta un factor de corrección, el cual corresponde a la mayor diferencia de las magnitudes de las tensiones de las barras que pertenecen al sistema, éste se define utilizando la siguiente expresión:

$$\beta = 1 - (\max(V_{max} - V_{min}))^2 \quad (2.7)$$

Por lo anterior, la formulación del índice propuesto por el autor es la relación entre la expresión (2.6) y (2.7) multiplicado por el voltaje de la barra en estudio V_i :

$$SVSI_i = \frac{\Delta V_i}{\beta * V_i} \quad (2.8)$$

2.1.2.2 Índice de estabilidad de línea

a. Índice rápido de estabilidad de tensión FVSI [Fast Voltage Stability Index] (Moghavvemi & Faruque, 1999)

Para la formulación de este índice, se hace uso de la ecuación cuadrática de tensión, que se obtiene de la representación circuital de un sistema de carga conectado a un barraje infinito. Una estimación, para el cálculo de este índice, es no considerar la diferencia angular entre las tensiones de las barras a las que está conectada la línea. Por lo tanto, se obtiene la expresión (2.9)

$$FVSI = \frac{4 * Z^2 * Q_r}{V_s^2 * X} \quad (2.9)$$

b. Índice de estabilidad de línea Lmn (Moghavvemi & Faruque, Real-time contingency evaluation and ranking technique, 1998)

El índice propuesto por los autores, se desarrolla bajo la idea del concepto de flujo de potencia a través de una línea, como la presentada en la Figura 2.1 bajo la expresión (2.10).

$$L_{mn} = \frac{4 \times Q_r}{(V_s \sin(\theta - \delta))^2} \quad (2.10)$$

c. Factor de estabilidad de línea LQP (Moghavvemi & Faruque, Technique for assessment of voltage stability in ill-conditioned radial distribution network, 2001)

Para la formulación de este índice, se considera el efecto de la potencia reactiva de forma indirecta al considerar la tensión en el nodo de envío. El índice se calcula bajo la expresión mostrada en (2.11).

$$LVSI = \frac{4 * r * P_r}{(V_s * \cos(\theta - \delta))^2} \quad (2.11)$$

2.2 Minería de datos

La minería de datos consiste en la aplicación de técnicas para analizar bases de datos, y mediante algoritmos de aprendizaje de máquina, obtener patrones o comportamientos y de esta forma, llegar a un nivel valido para garantizar el conocimiento (Tou JT, 1997).

2.2.1 Algoritmo de agrupación *k means*

El agrupamiento pertenece al aprendizaje no supervisado en la metodología del aprendizaje de máquina. Consiste en dividir una base de datos, definida con ciertos parámetros, en diferentes grupos. Estos grupos tienen como característica, una alta similitud de los datos que pertenecen a él y una gran diferencia con los datos que pertenecen a un grupo diferente (Tou JT, 1997).

Uno de los algoritmos de agrupamiento más conocidos y utilizados es el *k* medias o de su traducción al inglés *k means* (MacQueen, 1968). Este algoritmo, subdivide la base de datos en un número *k* de grupos, definido por el usuario. La pertenencia de un dato a un grupo, se define por las características de similitud que presente con el valor promedio del grupo. El valor medio de cada grupo es llamado centroide. Este valor, representa adecuadamente a los datos que pertenecen a un mismo grupo.

Este es un proceso iterativo en el cual, la definición de las características similares entre los datos se evalúa utilizando una distancia, esta se calcula de diferentes formas. En la Figura

2.2 se muestra la agrupación de datos en dos dimensiones utilizando el algoritmo propuesto.

Figura 2.2: Metodología de agrupación en dos dimensiones



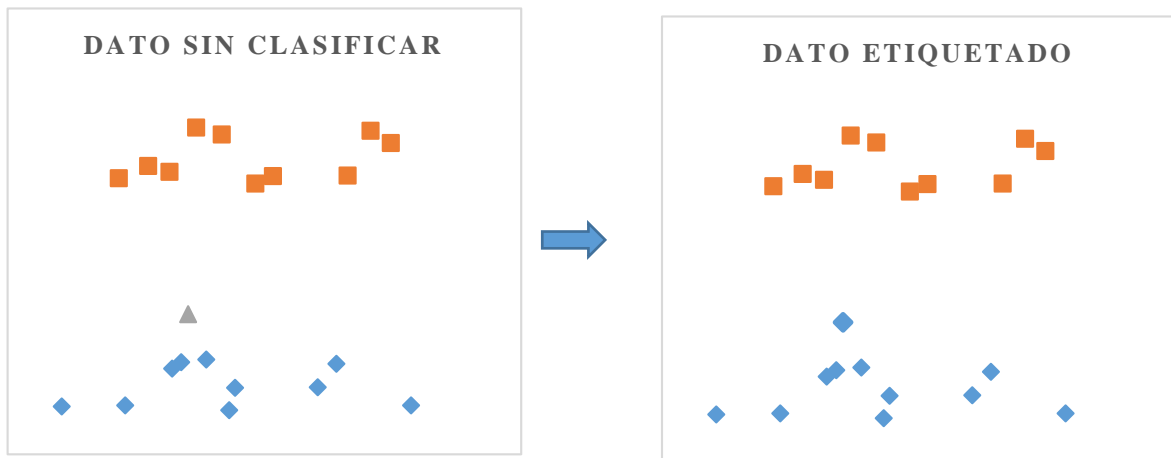
2.2.2 Algoritmo de clasificación k vecinos más cercanos

El proceso de clasificación pertenece a la metodología de aprendizaje de máquina, el cual consiste en darle una etiqueta a un dato nuevo, de acuerdo, a sus características o comportamiento (Moreno, 2004).

El algoritmo k vecinos más cercanos, knn por su traducción al inglés (k nearest neighbors), es un algoritmo de fácil implementación (Fix & Hodges, 1970). La metodología del algoritmo consiste en determinar la etiqueta de un dato, basado en la clase con mayor frecuencia, a la que pertenecen sus vecinos más cercanos (Moujahid & Inza I, 2008) (Russell & Norvig, 2003). En este algoritmo, se calcula la distancia del dato con etiqueta desconocida, a cada uno de los datos de entrenamiento. De acuerdo a las clases de los k vecinos más cercanos se asigna una clase al nuevo dato.

Por lo anterior, la técnica de clasificación del algoritmo knn , consta de una etapa de entrenamiento y una etapa de aplicación. En la etapa de entrenamiento, se busca el valor adecuado de la cantidad k de datos para la selección del vecindario de decisión y la forma de calcular la distancia a los datos. Los datos que pertenecen al vecindario, se consideran como los vecinos más cercanos. En la etapa de aplicación, se ejecuta el algoritmo con los parámetros adecuados para etiquetar el dato.

En la Figura 2.3 se muestra la metodología de aplicación del algoritmo en un espacio bidimensional, cuando la base de datos de entrenamiento solo presenta dos clases.

Figura 2.3: Metodología de clasificación en dos dimensiones

Una forma de mejorar la metodología propuesta de este algoritmo, es considerar diferentes modificaciones, para así asegurar que la asignación de clase del nuevo dato se realiza adecuadamente. Las posibles modificaciones son:

a. *knn* con distancia media

Esta modificación considera asignar la clase del nuevo dato, de acuerdo al mínimo de la distancia promedio de las clases cercanas.

b. *knn* con ponderación de pesos

Utilizando una función que tenga un comportamiento inverso de la distancia, se busca que los valores de los datos más cercanos, tengan mayor peso. Como resultado, la asignación de clase se realiza de acuerdo a la suma de mayor valor de los vecinos más cercanos (Wang, Zeng, & Yang, 2008).

c. *knn* con rechazo

El *knn* con rechazo, establece más restricciones para que el dato a clasificar, cumpla ciertas garantías, sin embargo, con esta modificación el dato no es clasificado si no cumple estas restricciones.

d. *knn* con pesado de variable

El algoritmo *knn* con pesado de variable, involucra un cambio en la forma para calcular la distancia entre los datos. Por lo tanto, implica que algunas variables tienen más importancia que otras. Para realizar la clasificación mediante este método, se debe encontrar un factor de pesos que pondere adecuadamente cada una de las variables.

2.2.3 Formas para calcular la distancia

Una forma de analizar las características que existen entre diferentes datos, se realiza calculando la distancia que exista entre ellos. Por lo tanto, hay diversas formas de encontrar la distancia entre datos. Entre las formas más relevantes se encuentran la Norma-1 o distancia Manhattan $\|d\|_1$, Norma-2 o distancia Euclídea $\|d\|_2$, Norma- L_∞ o distancia Chebyshev $\|d\|_\infty$ y Norma Euclídea ponderada $\|d\|_w$ (Esqueda Elizondo & Palafox Maestre, 2005).

El cálculo de la distancia que existe entre dos puntos X_k y X_m , por medio de las diferentes formas, sin importar el espacio dimensional \mathbf{R}^n (n dimensiones) al que ambos pertenezcan, se realiza mediante las expresiones (2.12) hasta (2.15).

$$\text{Distancia Manhattan} \quad \|d\|_1 = \sum_{i=1}^n |X_k - X_m|_i \quad (2.12)$$

$$\text{Distancia Euclídea} \quad \|d\|_2 = \sqrt{\sum_{i=1}^n (X_k - X_m)_i^2} \quad (2.13)$$

$$\text{Distancia Chebyshev} \quad \|d\|_\infty = \max(|X_k - X_m|_i) \quad (2.14)$$

$$\text{Distancia Euclídea ponderada} \quad \|d\|_w = \sqrt{(X_k - X_m)^T W (X_k - X_m)} \quad (2.15)$$

W matriz simétrica positiva que describe la ponderación de cada variable

3. Metodología para determinación de áreas críticas de estabilidad de tensión utilizando aprendizaje de máquina

3.1 Introducción

En este capítulo se presenta la metodología propuesta para determinar áreas críticas de estabilidad de tensión. Se utilizan herramientas de simulación de sistemas de potencia, para la generación de bases de datos, técnicas de agrupamiento de datos para definir las áreas críticas y finalmente, técnicas de aprendizaje supervisado, para la identificación de las áreas críticas, ante un evento determinado.

3.2 Modelado y simulación del sistema eléctrico de potencia

El comportamiento de los sistemas de potencia se puede estudiar principalmente de dos formas, con modelos estáticos y con modelos dinámicos. Para el caso de estabilidad de tensión, considerando las dinámicas lentas de los elementos, los modelos estáticos son una representación adecuada del sistema de potencia. (Amraee, Ranjbar, Mozafari, & Sadati, 2007).

A partir de lo anterior, los elementos de los sistemas de potencia utilizados en esta tesis, no consideran la dinámica asociada a su operación; por tanto, los elementos de regulación de velocidad o transitorios de los generadores, no son considerados para este modelo.

Para la aplicación de la metodología, en esta tesis se utiliza el sistema IEEE de 14 barras, el cual por su topología, permite el análisis de estabilidad de tensión. Este sistema será descrito en detalle en el capítulo 4.

La propuesta metodológica considera un software que permite la simulación de sistemas de potencia, en el cual se solucionan flujos de potencia; con los valores obtenidos de la variable de estado de tensión para cada nodo y la topología del sistema, se calculan los índices descritos en el capítulo 2. Para desarrollar la simulación del sistema eléctrico de potencia se utiliza el toolbox PSAT (Milano, 2001) de Matlab. Este software facilita el proceso de simulación, para sistemas de potencia con cambios en las condiciones de operación del sistema, para la obtención de una base de datos con validez suficiente para la implementación de las técnicas de aprendizaje de máquina propuestas.

3.3 Generación base de datos

En esta propuesta se requiere una base de datos con muchas condiciones de operación del sistema eléctrico, para implementar una técnica de agrupación. Para la generación de esta base de datos, se utiliza una matriz de variación de carga que represente adecuadamente todas las posibles condiciones de demanda y generación del sistema; adicionalmente, se consideran contingencias (N-1) de líneas de transmisión y generadores, para representar completamente las posibles condiciones de operación del sistema. Después de obtenida la matriz de carga, se implementa un modelo de simulación que considere los cambios dados por la matriz de carga y mediante una estrategia basada en valores promedio, se compensa la condición de generación dada para cada condición.

Con los resultados obtenidos del proceso de simulación, se calculan los índices de estabilidad de tensión propuestos; éstos se utilizan en el proceso de reconocimientos de áreas inestables en tensión, como base para una futura clasificación de una nueva condición de operación del sistema.

3.3.1 Latin Hypercube

La base de datos debe contener información suficiente y relevante para obtener patrones del comportamiento de un sistema eléctrico de potencia. En esta tesis, se utiliza una técnica que permite obtener una matriz de carga homogénea, y así, una descripción adecuada de posibles condiciones que se pueden presentar en el sistema de potencia.

La técnica escogida para la creación de la matriz de carga homogénea es la técnica de muestreo Latin Hypercube (Helton & Davis, 2003). Esta técnica permite una densificación adecuada de las variables de cada condición, cuando se tiene un tamaño de muestra pequeña; sin embargo, mantiene las condiciones requeridas para realizar un proceso de muestreo aleatorio.

Para realizar el muestreo utilizando esta técnica, se considera una función definida de probabilidad que se aplica a un vector \mathbf{x} que tiene n componentes, como se muestra en (3.1)

$$q = f(\mathbf{x}) \quad \text{y} \quad \mathbf{x} = [x_1, x_2, \dots, x_n] \quad (3.1)$$

Con la definición del rango de variación para cada variable x_j , ésta se divide en h intervalos contiguos con igual probabilidad. Esta probabilidad puede ser descrita de forma subjetiva o empírica y estar especificada bajo algunas restricciones. Así, se escogen aleatoriamente valores para x de acuerdo al rango y la distribución de probabilidad. Por lo tanto, se obtiene una matriz con factores de ponderación para cada variable que pertenece a x , tal como se muestra en (3.2).

$$x_i = [x_{i1}, x_{i2}, \dots, x_{in}], \quad i = 1, 2, \dots, h \quad (3.2)$$

La matriz de carga obtenida es diferente para cada sistema de potencia y depende de la cantidad de variables involucradas. A manera de ejemplo, para el sistema de potencia IEEE de 14 barras, se encuentra una matriz de variación de carga de 12000 condiciones de incremento de carga. Para compensar cada condición de carga, las barras donde se encuentra la generación también se alteran por el valor promedio del factor ponderador para cada caso t de carga. Una representación de la matriz de carga se muestra en (3.3), en donde cada barra de carga se afecta por un incremento.

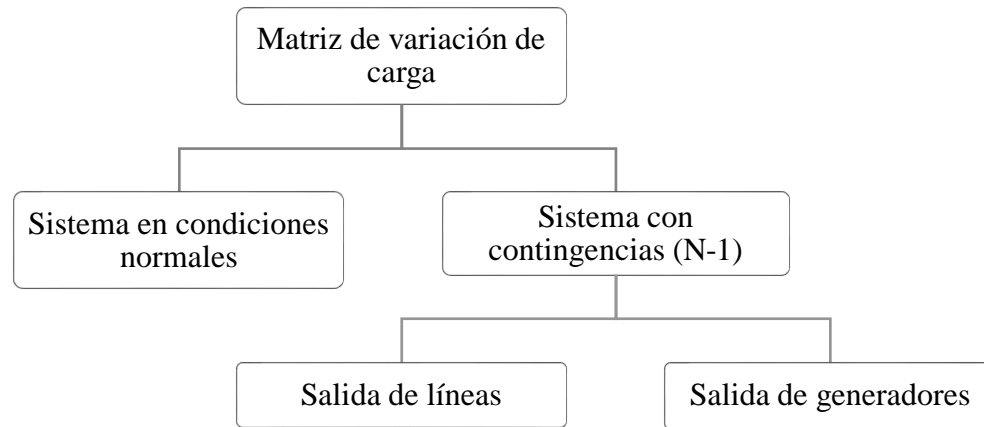
$$\begin{array}{rccccc}
 & \text{barra 1} & \text{barra 2} & \cdots & \text{ultima barra de carga} \\
 \text{caso 1} & \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\
 \text{caso 2} = & \alpha_{21} & \alpha_{22} & \cdots & \alpha_{1N} \\
 \vdots & \vdots & \vdots & \vdots & \vdots \\
 \text{caso } t & \alpha_{t1} & \alpha_{t2} & \cdots & \alpha_{tN}
 \end{array} \quad (3.3)$$

3.3.2 Solución del flujo de carga

Una vez se obtiene la matriz de carga, se simula cada condición de operación, según la matriz de casos que se muestra en (3.3). Adicionalmente, se consideran contingencias (N-1) del sistema de potencia, para cada condición de demanda obtenida utilizando la técnica de muestreo Latin Hypercube. Por lo tanto, para cada contingencia del sistema (salida de línea o de generador), se simulan flujos de carga, utilizando el toolbox PSAT, para obtener los valores de la variable de estado, y con todas estas condiciones obtener una mejor descripción del comportamiento del sistema de potencia bajo análisis. Este proceso se esquematiza en la Figura 3.1, y de esta forma se describen completamente las condiciones de operación de un sistema de potencia eléctrico.

Para la consideración del sistema eléctrico de potencia sin contingencias, se ejecuta el flujo de carga, con la matriz de condiciones de carga obtenida. Para el caso de las contingencias con líneas, se considera la desconexión de la línea n y se ejecuta el flujo de carga para cada una de las condiciones de la matriz de carga; después se reconecta la línea y se considera la desconexión de otra línea; el proceso se repite hasta analizar todas las líneas de transmisión que componen el sistema eléctrico de potencia. El procedimiento para analizar las contingencias de los generadores es igual al presentado para las líneas.

Figura 3.1: Esquema de condiciones de operación para un sistema de potencia.



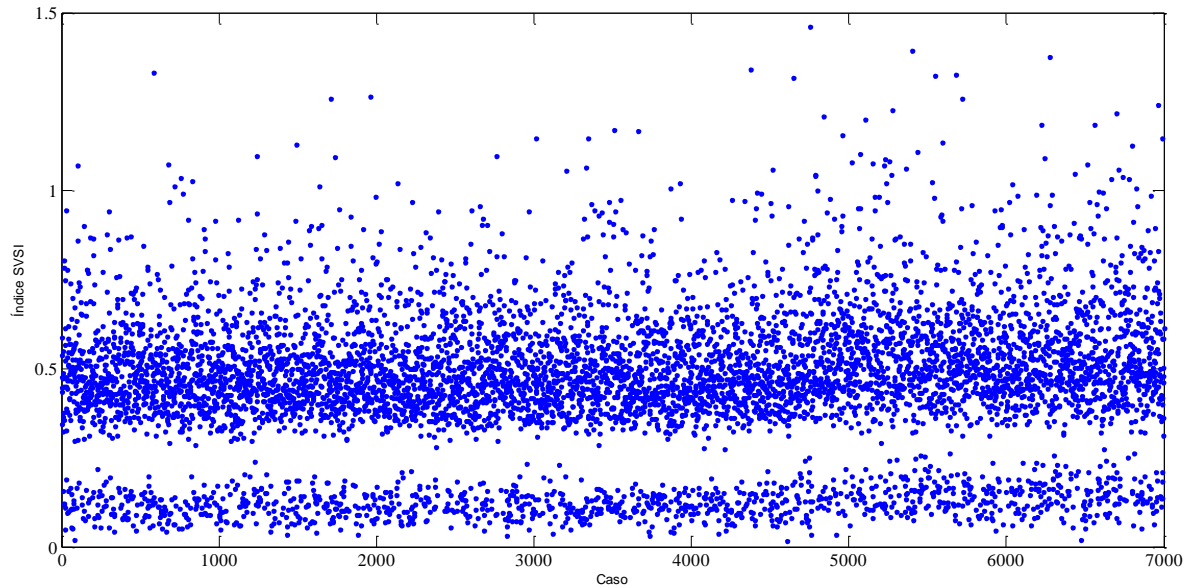
Como ejemplo, para el sistema considerado en esta tesis, se tienen 312000 casos de estudio en las cuales se consideran 12000 condiciones de carga para cada estado del sistema (sin contingencia y con contingencia (N-1) para líneas y generadores)

3.3.3 Índices de estabilidad de tensión

Con los resultados del flujo de carga, para cada uno de los casos de estudio o condiciones de operación del sistema eléctrico de potencia, se determinan los índices de estabilidad de tensión de barra y de línea presentados en el capítulo 2. Por lo tanto, la base de datos se conforma con los resultados de los flujos de carga obtenidos por el toolbox PSAT, y los índices de barra VCPI, índice L e índice SVSI y los índices de línea Lmn, LQP y FVSI.

De acuerdo al valor de los índices, para cada una de las condiciones de operación, se pueden obtener criterios de estabilidad o de inestabilidad. En la Figura 3.2 se muestra el comportamiento del índice SVSI para algunas condiciones de operación, a manera de ejemplo para la barra 12.

Figura 3.2: Comportamiento del índice SVSI para la barra 12 considerando algunas condiciones de operación.



3.3.4 Descriptores utilizados y almacenamiento de la base de datos

Después del cálculo de cada uno de los índices de estabilidad de tensión para cada una de las condiciones de operación o casos de estudio, se guardan estos resultados ordenadamente para generar la base de datos que se utiliza en el proceso de *agrupamiento* y el entrenamiento y parametrización de la técnica de clasificación.

Así, para cada condición de operación simulada, se obtienen las diferentes variables y se almacenan como descriptores, los cuales contienen información suficiente para entender el comportamiento de cada elemento que pertenece a un sistema de potencia bajo determinada condición de carga.

Los descriptores utilizados para cada una de las condiciones de operación son: La magnitud y ángulo de tensión para cada una de las barras que conforman el sistema, los flujos de potencia a través de todas las líneas de transmisión que interconectan los barrajes del sistema, la potencia activa y reactiva generada por cada uno de los generadores y los valores de los índices de barra y línea. La Tabla 3.1 muestra los descriptores utilizados para cada una de las condiciones de operación.

La utilización de todos los descriptores para este problema, se fundamenta en la complementariedad de los índices de línea y de barra y los valores de la tensión en cada barra, la potencia generada y los flujos de potencia.

Tabla 3.1. Descriptores utilizados para describir cada condición de operación.

Condición de operación	Tensión		Índices de nodo			Índices de línea			Potencia generada		Flujo de potencia por las líneas de transmisión.			
	$ V $	Θ	VCPI	L	SVSI	Lmn	LQP	FVSI	P_g	Q_g				
X_k	(1-m)	(1-m)	(1-m)	(1-m)	(1-m)	(1-n)	(1-n)	(1-n)	(1-g)	(1-g)	P_{ij}	P_{ji}	Q_{ij}	Q_{ji}

Donde k corresponde a la condición de operación, m corresponde a la cantidad de barras del sistema, n corresponde a la cantidad de líneas que componen el sistema, g corresponde a la cantidad de generadores conectados al sistema, e i y j corresponden a los flujos de potencia por la línea de transmisión desde el nodo emisor hasta el nodo receptor.

Para conformar la base de datos, se deben considerar las variables que pueden alterar las técnicas de agrupación y clasificación. A manera de ejemplo, la tensión del nodo slack, no aporta información importante dentro de la técnica de agrupación, ya que para cada condición de operación, esta tensión no varía; por lo tanto, este descriptor debe ser excluido de la base de datos para evitar inconsistencias en de la técnica propuesta en esta tesis.

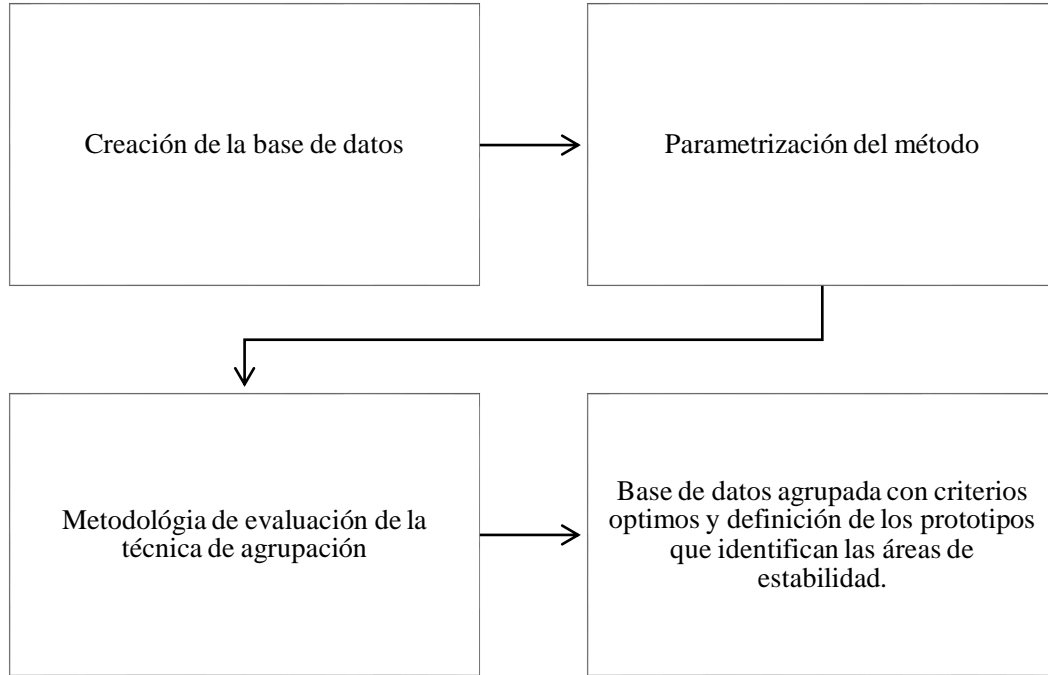
3.4 Técnica de agrupación basada en el método *k means*

La técnica de agrupación basada en el método *k means*, considera la distancia entre dos puntos para evaluar la similitud entre ellos, y así representar esta característica utilizando el valor medio, conocido como centroide del grupo. Para este caso, la característica de similitud está dada por los índices de estabilidad de tensión de línea y de barra, los valores de las tensiones de las barras, los flujos de potencia por las líneas de transmisión y la potencia generada.

Cada grupo formado mediante la técnica de agrupación, está caracterizado por un centroide o prototipo, para eventualmente caracterizar las áreas críticas de estabilidad.

Para ejecutar la técnica de agrupación, se deben definir las variables de entrada, la parametrización del algoritmo y un método de evaluación, para así obtener los valores adecuados de grupos para la base de datos utilizada. La parametrización del método está sujeta a la cantidad de condiciones de operación que conforman la base de datos y los descriptores utilizados. En la Figura 3.3 se ilustra un esquema de la técnica propuesta.

Figura 3.3: Esquema de agrupación mediante la técnica *k means*.



a. Descriptores utilizados y condiciones de operación

Para definir las variables de entrada del método, se debe elegir inicialmente una base de datos organizada y coherente; la base de datos, contiene la respuesta del sistema eléctrico de potencia ante alguna condición de carga o contingencia, así, cada vector que pertenece a la base de datos, contiene los descriptores suficientes y necesarios para iniciar el tratamiento de los datos. Estos vectores se definen como se muestra en la expresión (3.4)

$$x_i = [x_{i1}, x_{i2}, \dots, x_{ik}, \dots, x_{in}] \quad (3.4)$$

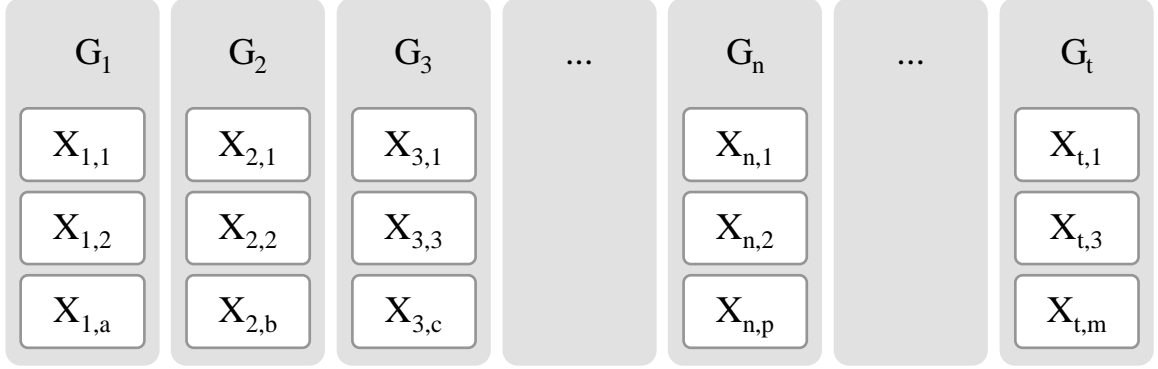
A partir de los valores de los índices de barra y de línea, se pueden definir criterios para evaluar la estabilidad que presenta el sistema de potencia ante cierta condición de operación.

b. Conformación de grupos de condiciones de operación

Uno de los parámetros que se debe definir para ejecutar la técnica de agrupación mediante la técnica *k means*, es la cantidad de grupos, para posteriormente evaluar la similitud de los datos. Así, un conjunto de grupos o un agrupamiento A_y , que se realice a una base de datos, está compuesto de t grupos totales. Cada uno de los grupos conformado, contiene m condiciones de operación del sistema eléctrico de potencia, y así el conjunto de las m_1 condiciones de operación que pertenecen a determinado grupo t_1 , poseen características similares entre ellas y características diferentes a las condiciones de operación m_2 que pertenecen al grupo t_2 .

Para cada agrupamiento A_y se utiliza la técnica *k means* para definir la cantidad de condiciones de operación m que están contenidas en un grupo t . Como resultado, todas las condiciones de operación que conforman la base de datos, están incluidas dentro de algún grupo, de acuerdo a sus características.

Figura 3.4: Agrupamiento realizado mediante la técnica *k means*.



Tomando como base el esquema mostrado en la Figura 3.4, se observa que todas las m condiciones de operación pertenecen a algún grupo, esta afirmación se representa mediante la expresión (3.5).

$$n = a + b + c + \dots + p + \dots m \quad (3.5)$$

Donde $a, b, c, \dots, p, \dots, m$ corresponden a la cantidad de condiciones de operación pertenecientes a cada grupo creado mediante la técnica de agrupación *k means*.

La pertenencia de una condición x_i a un grupo G_t determinado, está definido por los valores característicos de los valores de tensión en las barras, los valores de los índices de estabilidad de tensión de barra y de línea y en general, para todos los descriptores mostrados en la Tabla 3.1. Esto implica que las condiciones de operación que pertenece a determinado grupo son muy similares, tal y como se expresa en (3.6)

$$x_{t,1} \approx x_{t,2} \approx x_{t,3} \approx \dots \approx x_{t,m} \quad (3.6)$$

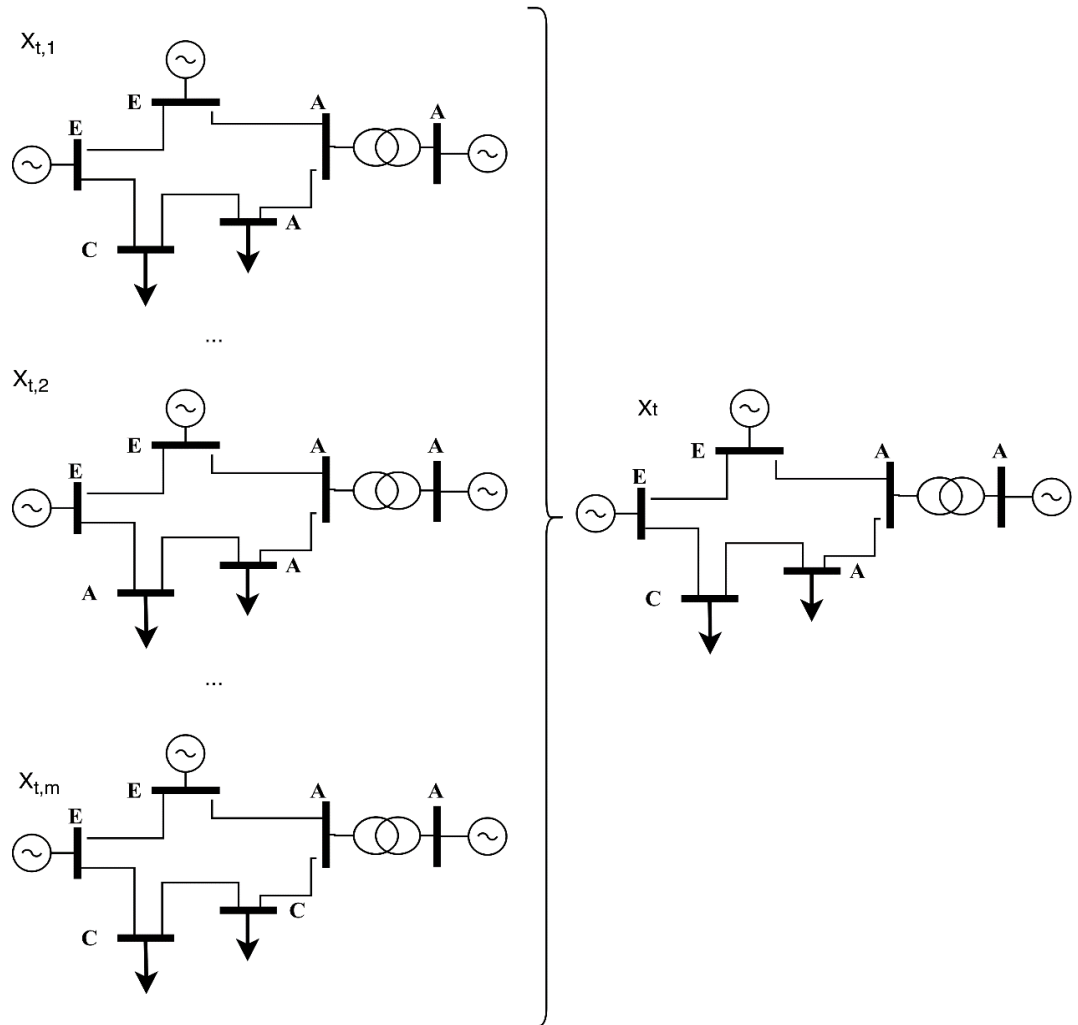
Por lo tanto, cada descriptor de la condición x_i , se parece al descriptor ubicado en la misma posición de la siguiente condición x_j , que pertenece al mismo grupo G_t .

$$x_i(x_{t,1}) \approx x_i(x_{t,2}) \approx x_i(x_{t,3}) \approx \dots \approx x_i(x_{t,m}) \quad (3.7)$$

De esta forma, es posible considerar que las m condiciones de operación, que pertenecen a determinado grupo G_i , se pueden representar mediante un prototipo; para el caso de la técnica de agrupación *k means*, es el centroide del grupo de las condiciones de operación que lo componen.

En la Figura 3.5 se presenta de forma gráfica un sistema de 6 nodos, y se muestran de forma cualitativa, los valores de los índices de barra para las condiciones de operación que pertenecen a un grupo G_i . Estos valores cualitativos están definidos en tres rangos: Alerta (A), crítico (C), estable (E). Así, cada uno de los nodos que pertenecen a este grupo, tiene un criterio de estabilidad y pueden ser representados por el prototipo propuesto.

Figura 3.5: Ejemplo de las condiciones de operación que conforman un grupo G_i y su prototipo representativo.



3.4.1 Estrategias para la evaluación del agrupamiento

La técnica *k means* permite obtener diferentes agrupamientos de acuerdo a los valores iniciales con los cuales se parametriza; por esta razón, se hace necesaria la aplicación de una metodología para determinar la calidad del agrupamiento que se realiza en una determinada base de datos. Así, para la evaluación del agrupamiento se definen tres criterios, uno que relaciona el promedio de los valores de las distancias al centroide y la distancia entre cada uno de los demás centroides creados, este criterio es una metodología matemática que ignora el sentido físico del problema que se requiere resolver mediante la agrupación; el segundo criterio involucra el sentido físico del problema, y relaciona el error ponderado de las tensiones de los nodos que pertenecen a determinada zona dentro de un grupo; el tercer y último criterio es un método cualitativo, en el cual se evalúa la relación que tiene un prototipo con otro, de acuerdo a los valores de los índices de estabilidad utilizados, los cuales son discretizados utilizando rangos de criticidad o zonas de severidad (Pérez Londoño, 2013).

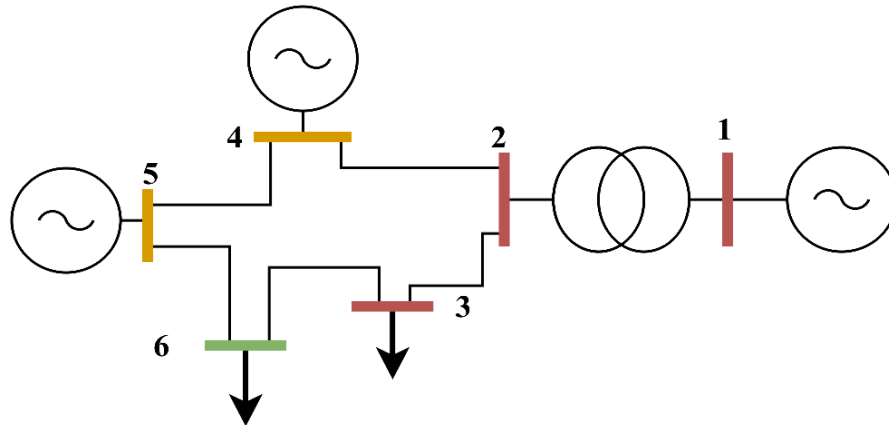
3.4.1.1 Definición de zonas de severidad

De acuerdo al valor numérico de los índices propuestos, para análisis de estabilidad de tensión, y de acuerdo a la experiencia de cada operador de red, se pueden definir ciertas zonas, en las cuales se podría clasificar este índice como un valor crítico o seguro, para asignar una representación cualitativa del valor del índice y permitir un mejor entendimiento de la condición de operación del sistema. Esta división cualitativa, permite indicar cuál es la mejor acción que un operador del sistema debe realizar ante determinada condición de operación del sistema eléctrico de potencia.

A manera de ejemplo, con la división en tres categorías de la estabilidad en tensión del sistema, tales como: Estable, alerta o inestable, se pueden mejorar las condiciones en que se muestran los resultados a los operadores del sistema, y las posibles acciones a realizar sobre el sistema. (Pérez Londoño, 2013). En las siguientes figuras, se muestran dos condiciones de operación, en donde se evidencia la división de las categorías con relación a los índices de tensión, en un sistema de potencia de prueba.

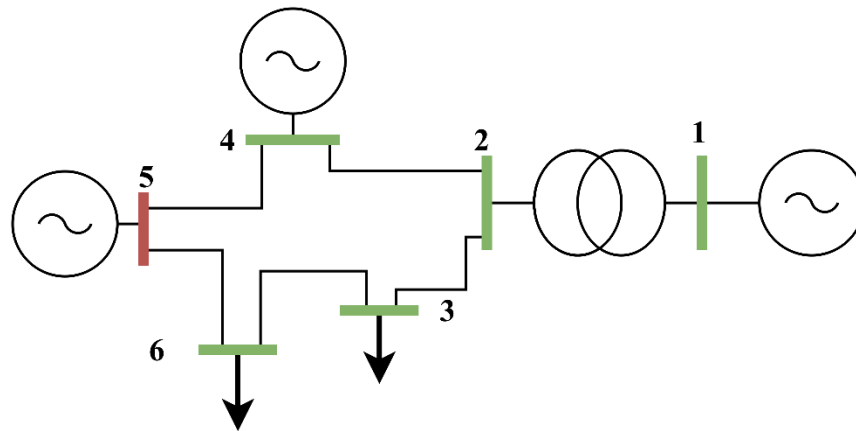
Como se puede observar en la Figura 3.6, se muestra que las barras 1, 2 y 3, presentan un valor crítico con respecto a la estabilidad de tensión y se debe realizar una entrega de reactivos en esta área para mejorar su perfil de tensión y evitar así un colapso; la barra 4 y 5 presentan una señal de alarma, debido a la cercanía eléctrica con las barras de valores críticos, por lo tanto, si no se ejerce una acción de control, estas pueden pasar a un estado de inestabilidad; el resto de barras del sistema se encuentran en valores estables de operación.

Figura 3.6: Condición de operación 1.



El mismo análisis es aplicable a la condición de operación mostrada en la Figura 3.7, en donde se observa que la barra 5, presenta un valor crítico de estabilidad y se deben enviar reactivos suficientes para evitar un colapso, mientras que las otras barras del sistema presentan valores estables de operación.

Figura 3.7: Condición de operación 2.



3.4.1.2 Estrategia de evaluación basada en el índice Davis Bouldin

Mediante el índice Davis Bouldin propuesto en (Davies & Bouldin, 1979), se determina la homogeneidad de cada agrupamiento, es decir, que tan compactos están los elementos que pertenecen a un grupo y la distancia de éstos con los demás datos que pertenecen a otro grupo diferente.

Por lo tanto, para un agrupamiento que contiene gt grupos totales, se puede definir un término $RS_{i,j}$, conocido como similaridad relativa entre un grupo G_i y otro G_j , utilizando la expresión (3.8)

$$RS_{i,j} = \frac{S_i - S_j}{l_{i,j}} \quad (3.8)$$

Donde S se calcula con base en la expresión (3.9), e indica la dispersión de las distancia de cada elemento i que pertenece a un grupo G_i , y $l_{i,j}$ es la distancia que existe entre el centroide G_i y el centroide G_j , tal y como se expresa en (3.10)

$$S_i = \frac{1}{mt_i} \sum_{i=1}^{mt} \|x_i - c_{G_i}\| \quad (3.9)$$

$$l_{i,j} = \|c_{G_i} - c_{G_j}\| \quad (3.10)$$

En las expresiones anteriores, x_i representa cada dato que pertenece al agrupamiento i y centroide C_{G_i} ; mt es la cantidad total de condiciones que pertenecen a ese grupo.

Así, después de definir la similaridad relativa para cada uno de los centroides formados por la técnica, se define la máxima similaridad relativa entre dos grupos $MRS_{i,j}$, mediante la expresión (3.11)

$$MRS_i = \max_{j=1,2,\dots,gt \ i \neq j} (RS_{i,j}), i = 1, 2, \dots, gt \quad (3.11)$$

De esta forma, el índice Davis Bouldin para un agrupamiento, se define como el valor promedio de los valores MRS_i encontrados, y así se define la calidad del agrupamiento con respecto a otro.

$$DB = \frac{1}{gt} \sum_{i=1}^{gt} MRS_i \quad (3.12)$$

3.4.1.3 Estrategia de evaluación basada en las diferencias de tensiones

Esta estrategia se propone en (Pérez Londoño, 2013) y ubica el problema de agrupación desde una perspectiva física, donde, se considera que un grupo de barras que tienen una cercanía geográfica, deben responder en igual medida a una perturbación.

Con base en la definición de zonas de severidad, para cada condición de operación se pueden determinar los elementos que pertenecen a cada zona Z_s de cada grupo k , creado mediante la técnica de agrupación *k means*. Así, se definen los descriptores que conforman la matriz de diferencias de tensión de barras, que pertenecen a una zona de severidad Z_s . Estos descriptores se muestran en la Tabla 3.2, donde nt representa la cantidad de nodos que pertenecen a la zona Z_s .

Tabla 3.2: Diferencias de tensión para los nt nodos que pertenecen a determinada zona de severidad Z_s para cada una de las condiciones de operación.

Índice de la diferencia	Diferencias posibles entre las tensiones de nodos que pertenecen a la zona				
1	$v_1 - v_2$	$v_1 - v_3$	$v_1 - v_3$...	$v_1 - v_{nt}$
2	$v_2 - v_3$	$v_2 - v_4$	$v_2 - v_5$...	
\vdots	\vdots	\vdots	\vdots		
$nt - 2$	$v_{nt-2} - v_{nt-1}$	$v_{nt-2} - v_{nt}$			
$nt - 1$	$v_{nt-1} - v_{nt}$				

La matriz de diferencias de tensión se debe calcular para cada una de las zonas de severidad. Con la expresión (3.13) se encuentra el error ponderado para cada una de las zonas de severidad.

$$\Delta V_{(Z_s)(x_{(t,m)})} = \frac{\sum_{i=1}^{nt} |(nt - i)v_i - \sum_{j=i+1}^{nt} v_j|}{\sum_{i=1}^{nt} (nt - i)} \quad \forall i \in Z_s \quad (3.13)$$

De este modo, se debe encontrar el error ponderado de diferencias de tensiones para cada condición de operación, en relación a la cantidad de zonas totales de severidad st . Por medio de la expresión (3.14), se calcula el error ponderado para cada condición de operación x_i .

$$\Delta V_{(x_{(t,m)})} = \frac{1}{st} \sum_{i=1}^{st} \Delta V_{(Z_i)(x_{(t,m)})} \quad (3.14)$$

Después, se calcula el error ponderado para el grupo G_t , que está conformado por las mt condiciones de operación.

$$\Delta V_{G_t} = \frac{1}{mt} \sum_{i=1}^{mt} \Delta V_{(x_{(t,i)})} \quad (3.15)$$

La definición de la calidad del agrupamiento formado por la técnica *k means*, se calcula mediante el error ponderado de la cantidad de grupos totales gt que se crean. A partir de la expresión (3.16) se obtiene el índice para cada agrupamiento.

$$\Delta V = \frac{1}{gt} \sum_{i=1}^{gt} \Delta V_{G_i} \quad (3.16)$$

3.4.1.4 Estrategia de evaluación mediante la definición de rangos cualitativos de los índices de estabilidad de tensión

Debido a la definición de zonas de severidad descrita en 3.4.1.1, se puede afirmar que los centroides creados mediante cualquier técnica de agrupación, tienen relación con un valor cualitativo, con respecto a los valores de los índices de estabilidad de tensión.

Así, se puede apreciar cualitativamente las características entre dos grupos diferentes G_i y G_j , para evaluar su relación de similitud. Por lo tanto, si dos centroides C_i y C_j presentan valores de índices diferentes, y al evaluar su valor cualitativo, por medio de la definición de zonas de severidad, se puede concluir que ambos grupos son iguales y como resultado, se redefinen como uno solo en un nuevo grupo G_n . Este nuevo grupo contiene todas las condiciones de operación m_i y m_j que pertenecen a los grupos anteriores.

En la Tabla 3.3, se presenta un sistema de 3 barras bajo condiciones de operación diferentes, con sus respectivos valores del índice SVSI y su definición cualitativa.

Tabla 3.3: Comparación entre dos centroides que pertenecen a dos grupos diferentes; al lado izquierdo se observan los valores obtenidos de los índices, y al lado derecho se observa su equivalente cualitativo.

Grupo 1		Grupo 2	
Barra	Índice	Barra	Índice
1	0.215	1	0.012
2	0.647	2	0.550
3	0.418	3	0.117

Grupo 1		Grupo 2	
Barra	Índice	Barra	Índice
1	E	1	E
2	A	2	A
3	E	3	E

En el ejemplo se puede observar, que aunque dos grupos sean clasificados como diferentes por el valor de sus índices de estabilidad de tensión, cuando se evalúa el criterio cualitativo de estos índices, resultan ser el mismo grupo, y así, representan la misma condición de operación del sistema, lo que a su vez ayuda a disminuir la cantidad de condiciones diferentes de operación, que se definen como prototipos representativos del sistema de potencia.

3.4.2 Determinación de áreas críticas de estabilidad

Con la metodología de agrupación y evaluación de la calidad del agrupamiento propuesto, se puede definir un algoritmo para definir con mayor exactitud y calidad la determinación de áreas críticas en relación con su estabilidad de tensión.

El diagrama de flujo que resume este algoritmo se muestra en la Figura 3.8. Mediante este diagrama de flujo, se representa esquemáticamente los pasos del algoritmo para determinación de áreas críticas.

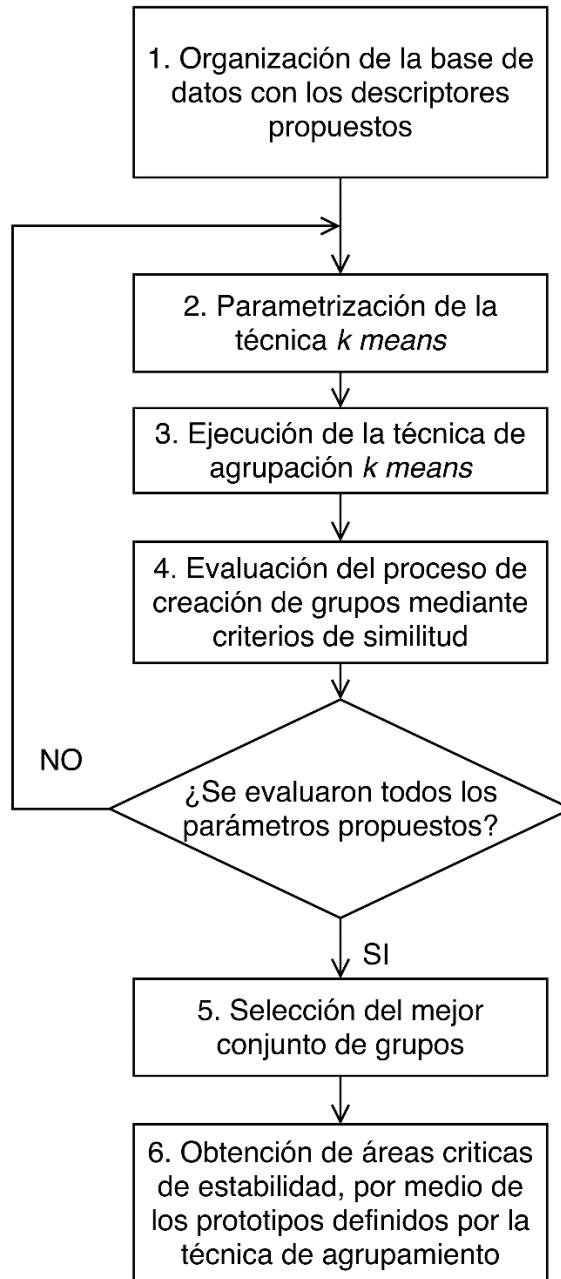
Cada uno de los pasos descrito en la Figura 3.8 se resume de la siguiente forma:

- a. La organización de la base de datos, y la definición de cada uno de los descriptores que representan cada una de las condiciones de operación establecidas, se definen para la entrada del agrupamiento.
- b. Para la parametrización del algoritmo *k means*, se especifica la cantidad de grupos que se van a crear y la norma para calcular la distancia entre los datos.
- c. Se ejecuta el algoritmo *k means*, con los parámetros especificados y la base de datos de entrada, para así, determinar las condiciones de operación que pertenecen a cada grupo.
- d. Mediante las técnicas de evaluación se calculan los índices Davis Bouldin, error de tensiones y rangos cualitativos para evaluar la calidad del agrupamiento.

Debido, a que para un mismo conjunto de datos, se pueden definir diferente cantidad de grupos y la forma de calcular la distancia entre los datos entonces, se repiten los pasos anteriores, para definir con seguridad los mejores criterios de parametrización.

Finalmente, a partir de los resultados de los índices de evaluación, se determina cuál es el mejor agrupamiento, que permite determinar los prototipos que mejor representan las diferentes condiciones de operación del sistema de potencia.

Figura 3.8: Diagrama de flujo para determinación de áreas críticas de estabilidad en tensión.

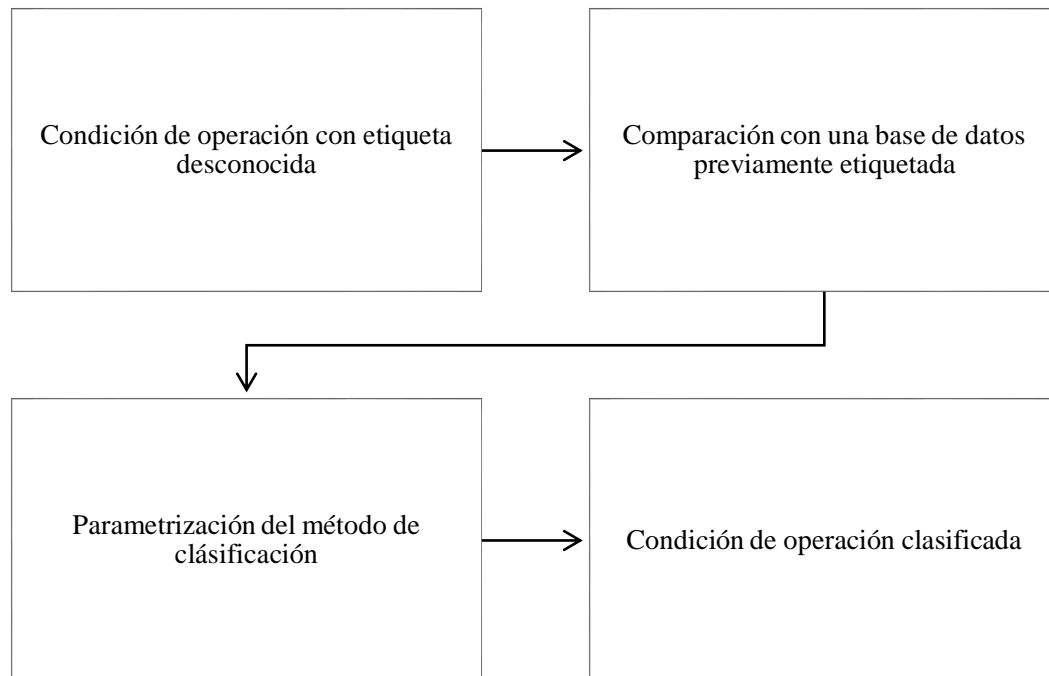


3.5 Técnica de clasificación basada en el método k vecinos más cercanos

La técnica de clasificación basada en el método k vecinos más cercanos knn , trabaja sobre la idea del aprendizaje de máquina supervisado, en el cual, se requiere etiquetar una nueva condición de operación, comparando con una base de datos que está completamente clasificada. De acuerdo a lo anterior, para el caso específico del sistema de potencia, se requiere evaluar la estabilidad de tensión de un área, cuando se presenta una nueva condición de operación (variación en la carga, contingencia, entre otras).

Según lo explicado, el algoritmo determina la distancia de la nueva condición de operación a cada una de las condiciones de operación previamente etiquetadas, mediante la técnica de agrupación k means, y clasifica la nueva condición de operación, seleccionando las etiquetas de las condiciones de operación que presenten menor distancia. En la Figura 3.9 se muestra un esquema de la técnica propuesta utilizando el algoritmo knn .

Figura 3.9: Esquema de clasificación mediante la técnica knn .



Debido a que ésta es una propuesta de clasificación en línea, se deben definir los mejores criterios para parametrizar el algoritmo knn , y así, ante una nueva condición de operación, garantizar un alto desempeño en el reconocimiento y su correcta clasificación.

3.5.1 Entrenamiento del algoritmo k vecinos más cercanos

Según el capítulo anterior, se debe definir criterios para ejecutar correctamente la técnica de knn . Esto implica definir la forma de calcular la distancia desde la nueva condición de operación a las demás condiciones que existen en la base de datos, especificar la cantidad de vecinos que se van a considerar para clasificar el nuevo dato y un método de evaluación de los parámetros especificados.

3.5.1.1 Modificaciones realizadas

El algoritmo knn calcula la distancia de un dato nuevo a cada uno de los datos contenidos en una base de datos, y de acuerdo a una vecindad definida, asigna la etiqueta de la mayoría de vecinos que están cerca. Por consiguiente, se debe definir la forma de calcular la distancia a los datos y la cantidad de vecinos que se consideran dentro del vecindario.

a. Forma de calcular la distancia

Como consecuencia de la metodología propuesta, la forma de calcular la distancia se debe evaluar de acuerdo a criterios que involucren físicamente el problema de estabilidad de tensión. Según los criterios definidos en la sección 0, la forma de calcular la distancia para el algoritmo de clasificación, debe estar relacionada con la forma utilizada en el proceso de agrupación de la base de datos inicial.

b. Selección de la cantidad vecinos a considerar para formar el vecindario

Uno de los parámetros de entrada de la técnica de clasificación, es la cantidad de datos que conformaran el vecindario de decisión. Para determinar la cantidad adecuada de vecinos, se deben idear estrategias de comparación y evaluación, que permitan obtener un mejor resultado para clasificar una nueva condición de operación, de la cual se desconoce su estabilidad en tensión.

c. Parametrización del algoritmo

Es recomendable evaluar diferentes criterios de parametrización del clasificador a utilizar; así, para determinado sistema de acuerdo a su topología, rangos de variación de carga en función del tiempo, elementos que lo componen, entre otras características, se pueden realizar algunas de las modificaciones propuestas en la sección 2.2.1. Para este tipo de problema, se prueban las modificaciones de ponderación de pesos utilizando funciones que varían de forma inversa a la distancia, distancia promedio y el algoritmo básico del knn .

3.5.1.2 Evaluación de la clasificación mediante validación cruzada

Debido a la dificultad para definir la cantidad de vecinos que se deben considerar, se propone como método de evaluación del clasificador, la técnica de validación cruzada (Gutierrez Gallego, Pérez Londoño, & Mora Flórez, 2010). Este método de evaluación, consiste en dividir una base de datos de la cual se conocen todas sus etiquetas y dividir las en t subgrupos, y evaluar el desempeño del clasificador, tomando como nueva base de datos

conocida a la conformada por todos los subconjuntos $\{S_1, S_2, S_3, \dots, S_{t-1}\}$ y utilizar el subconjunto S_t como el conjunto de datos de entrada con etiqueta desconocida.

El desempeño del clasificador propuesto, para uno de los subconjuntos S_t creados, se evalúa de acuerdo a la expresión (3.17).

$$desempeño_{(S_t)} = \frac{\text{cantidad de datos correctamente clasificados}}{\text{cantidad de datos}_{S_t}} \quad (3.17)$$

Así, el desempeño del clasificador total, es el promedio de los desempeños individuales de cada uno de los subconjuntos S_t tal y como se expresa en (3.8).

$$desempeño_{knn} = \frac{\sum_{i=1}^t \text{desempeño}_{S_i}}{t} \quad (3.18)$$

3.5.2 Parametrización del algoritmo k vecinos más cercanos y clasificación de una nueva condición de operación del sistema

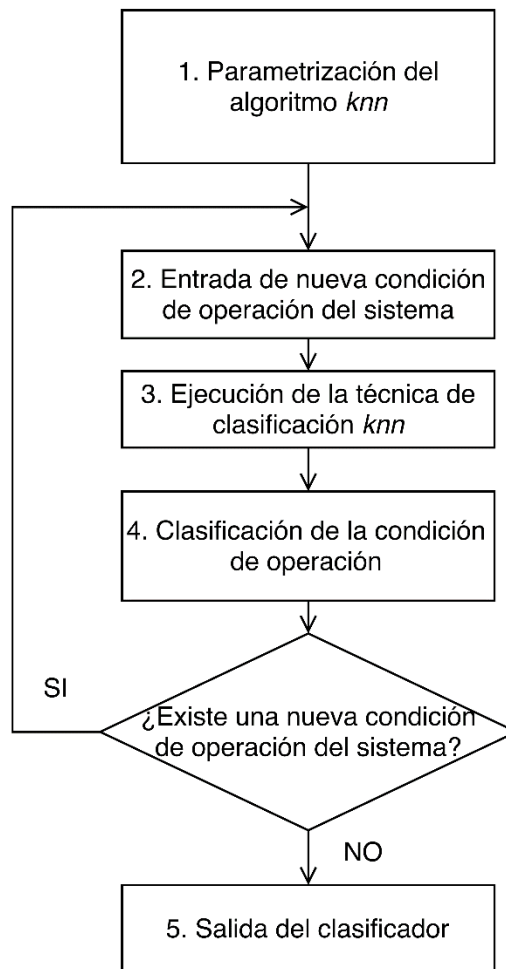
Como resultado de los valores obtenidos del desempeño del clasificador, y debido a la necesidad de definir los parámetros de entrada para ejecutar el algoritmo de clasificación, se seleccionan los parámetros de entrada del algoritmo knn , ante una nueva condición de operación desconocida, con base en el mejor rendimiento obtenido mediante la validación cruzada. Los parámetros a definir son: la forma de calcular la distancia (Norma 1, Norma 2, entre otras), la modificación al algoritmo básico (Distancia Media, Ponderación de la distancia, con rechazo, entre otras), y la cantidad k de vecinos que se van a considerar para formar el vecindario de decisión.

En la Figura 3.10 se muestra un esquema del proceso de clasificación de una nueva condición de operación del sistema

Cada uno de los pasos descrito en la Figura 3.10 se resume de la siguiente forma:

- a. Definición de un método de evaluación para parametrizar la técnica de clasificación.
- b. Cambio en la condición de operación del sistema y medición de las variables de estado.
- c. Proceso de clasificación mediante el algoritmo knn .
- d. Clasificación y análisis de la condición de operación.

Figura 3.10: Esquema de clasificación mediante la técnica *knn*



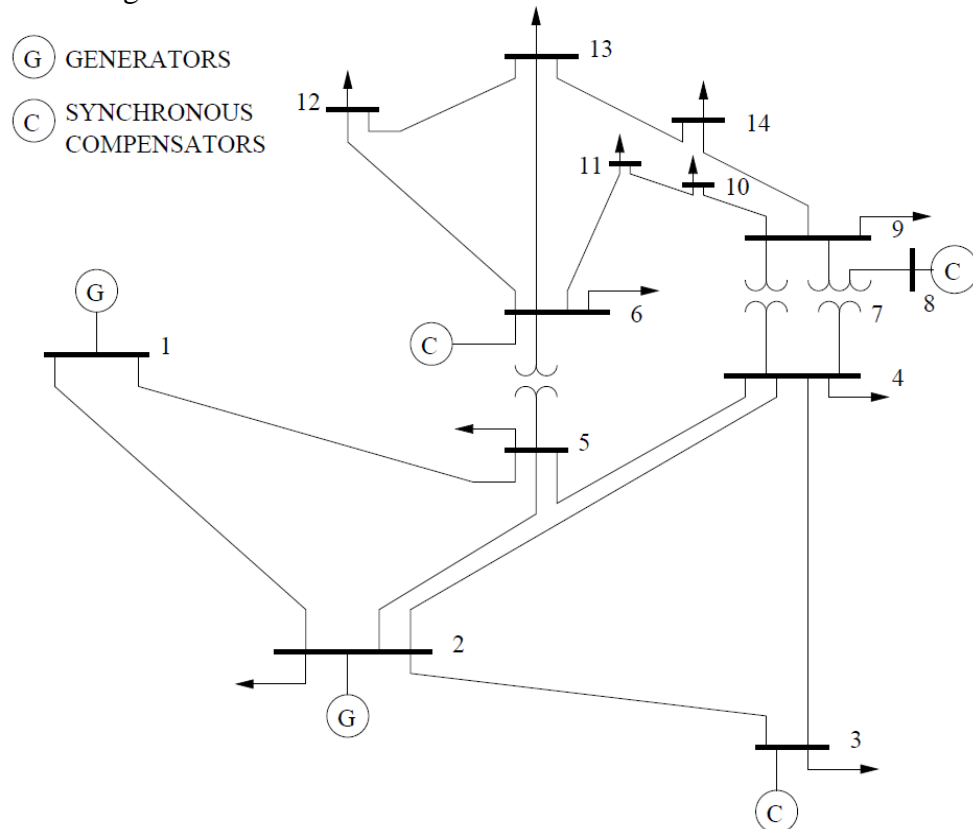
4. Aplicación de la metodología propuesta

En este capítulo, se presentan los resultados de la metodología propuesta para la identificación en línea de áreas con problemas de estabilidad de tensión, aplicada al sistema IEEE de 14 barras.

4.1 Descripción del sistema de prueba IEEE de 14 barras

Debido a la necesidad de validar la propuesta de identificación en línea, se propone aplicar esta metodología al sistema de potencia IEEE de 14 barras. Este sistema cuenta con 14 barras, nueve de ellos son nodos de carga y cinco son nodos que tienen conectados generadores síncronos, adicionalmente, el sistema cuenta con 20 líneas de transmisión. El sistema ha sido utilizado en diversos artículos por sus características topológicas y su aplicación en el análisis de la estabilidad de tensión (Nolasco Contreras, 2012). Un diagrama unifilar del sistema se muestra en la Figura 4.1 (Mena Kodsi & Cañizares, 2003).

Figura 4.1: Diagrama unifilar del sistema IEEE de 14 barras



La contingencia que presenta mayor impacto en este sistema, con relación a la estabilidad de tensión, es la desconexión de la línea que conecta los barrajes 1 y 2, ya que implica la desconexión de uno de sus generadores con mayor capacidad.

Debido a que este es un sistema pequeño, tiene bien definidas las áreas de generación y demanda, lo cual ayuda para una evaluación de la metodología propuesta en el capítulo 3 de esta tesis.

4.2 Descripción de las condiciones de operación para la validación de la propuesta

Para la validación de la propuesta, se genera una base de datos compuesta de diferentes condiciones de operación, que represente adecuadamente los posibles escenarios que se presentan en un sistema de potencia. Por lo tanto, por medio de simulación se obtienen las diferentes condiciones de operación al variar de forma aleatoria la demanda en los nodos, desde 0.5 a 2.0 p.u.

4.2.1 Prueba en el sistema

Considerando el tamaño del sistema IEEE de 14 barras, se obtiene una matriz de variación de carga, utilizando la técnica de muestreo Latin Hypercube. Esta matriz de carga contiene 12000 escenarios de variación para los 14 nodos del sistema. Adicionalmente, se analizan condiciones especiales de operación, en este caso, contingencia N-1, donde se retira cada una de las líneas de transmisión, se realiza la variación de la carga, y se vuelve a reconectar la línea para repetir el proceso con todas las líneas del sistema y generadores disponibles. Como consecuencia, se obtienen 312000 condiciones de operación del sistema eléctrico, de las cuales, al realizar el flujo de carga se obtiene la convergencia para 259701 casos, debido a las restricciones del sistema.

4.2.2 Descriptores utilizados en la base de datos

Con el resultado obtenido para el flujo de carga, con el toolbox PSAT, y la matriz de condiciones de carga y considerando las contingencias, se obtiene una base de datos compuesta por 259701 casos. Cada uno de estos casos está caracterizado por 220 descriptores que debido a las consideraciones del método de agrupación, se eliminan los valores que no aportan condiciones apropiadas. Por consiguiente, se eliminan los descriptores de tensión del nodo slack y los índices de estabilidad relacionados con ese nodo.

En total, se obtiene una matriz de 259701 filas por 213 columnas como la base de datos de entrada para evaluar la metodología propuesta.

4.3 Determinación de las áreas críticas de estabilidad

La aplicación de la metodología para determinación de áreas críticas de estabilidad al sistema IEEE de 14 barras se evalúa mediante los pasos descritos en la Figura 3.8.

4.3.1 Matriz de condiciones de operación

La matriz de condiciones de operación de entrada para este sistema consta de 259701 casos, obtenida mediante la técnica de muestreo Latin Hypercube para la variación de la carga y la aplicación de esta variación cuando el sistema se somete a diferentes contingencias. Sin embargo, debido a la necesidad de obtener las áreas más representativas, y como método para depurar la base de datos antes de utilizar la técnica de agrupamiento, se seleccionan aquellas condiciones de operación, en donde al menos uno de los índices de línea LQP o el índice de barra SVSI, se encuentra catalogado como crítico, según la definición de zonas de severidad. Así, se obtienen 32265 condiciones de operación que representan este comportamiento.

4.3.2 Selección de los parámetros de la técnica de agrupamiento

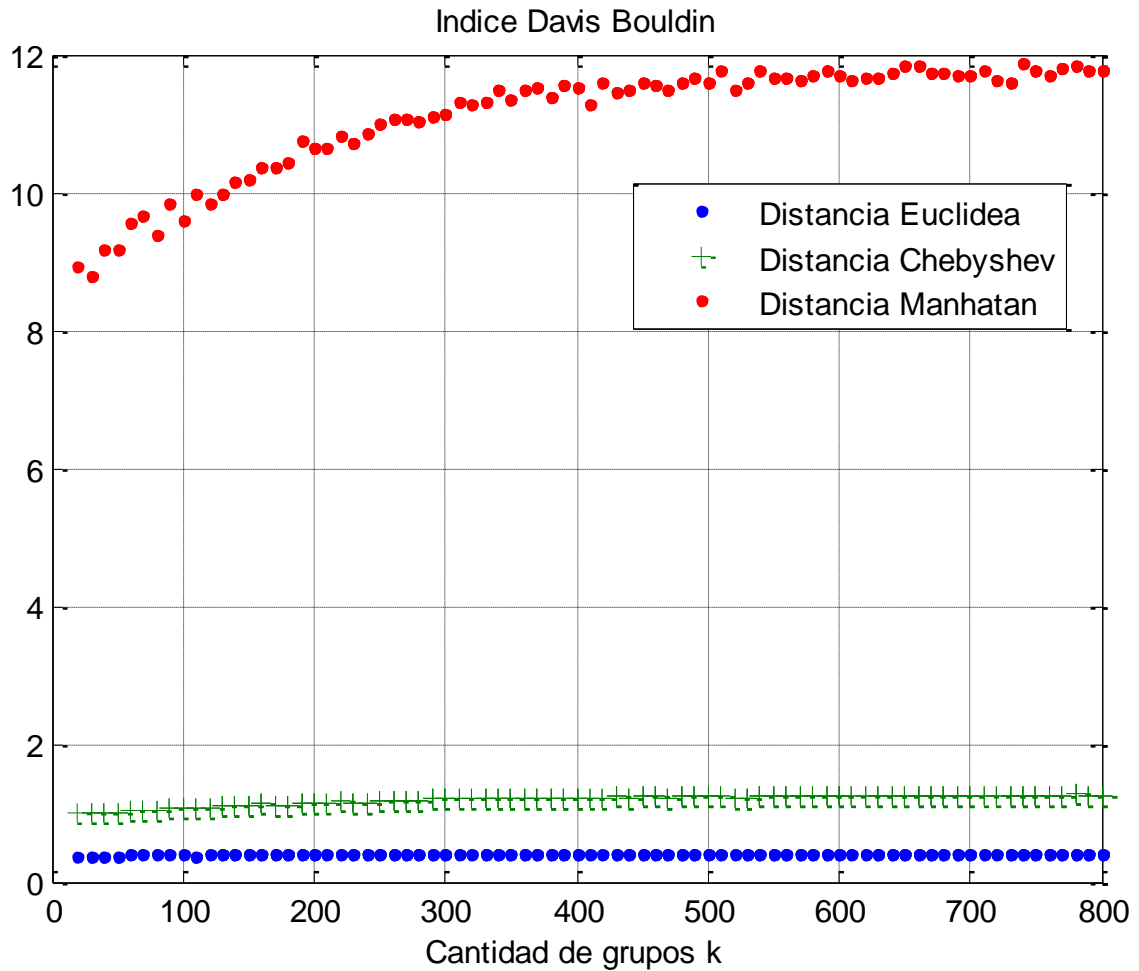
Para determinar los parámetros de la técnica de agrupamiento, se realiza una prueba extensiva, que consiste en variar la cantidad de grupos que se crean al aplicar el algoritmo *k means*, desde 20 hasta 800. Esta estrategia sirve para determinar, cuál es la forma adecuada para calcular la distancia entre dos puntos, y la cantidad de prototipos que mejor representan las diferentes condiciones de operación del sistema.

4.3.2.1 Selección de la norma

Para determinar la forma de calcular la distancia, se realiza la prueba agrupando la base de datos de entrada, y de acuerdo a los resultados del índice Davis Bouldin para cada agrupamiento, se selecciona la mejor norma. En la Figura 4.2 se muestra el resultado del índice para cada uno de los agrupamientos.

A partir de los resultados, se observa que la Norma Euclídea es la que presenta menor índice de error; por lo tanto, es la que permite la adecuada distribución de los centroides.

Figura 4.2: Índice Davis Bouldin para cada uno de los agrupamientos variando la forma de calcular la distancia



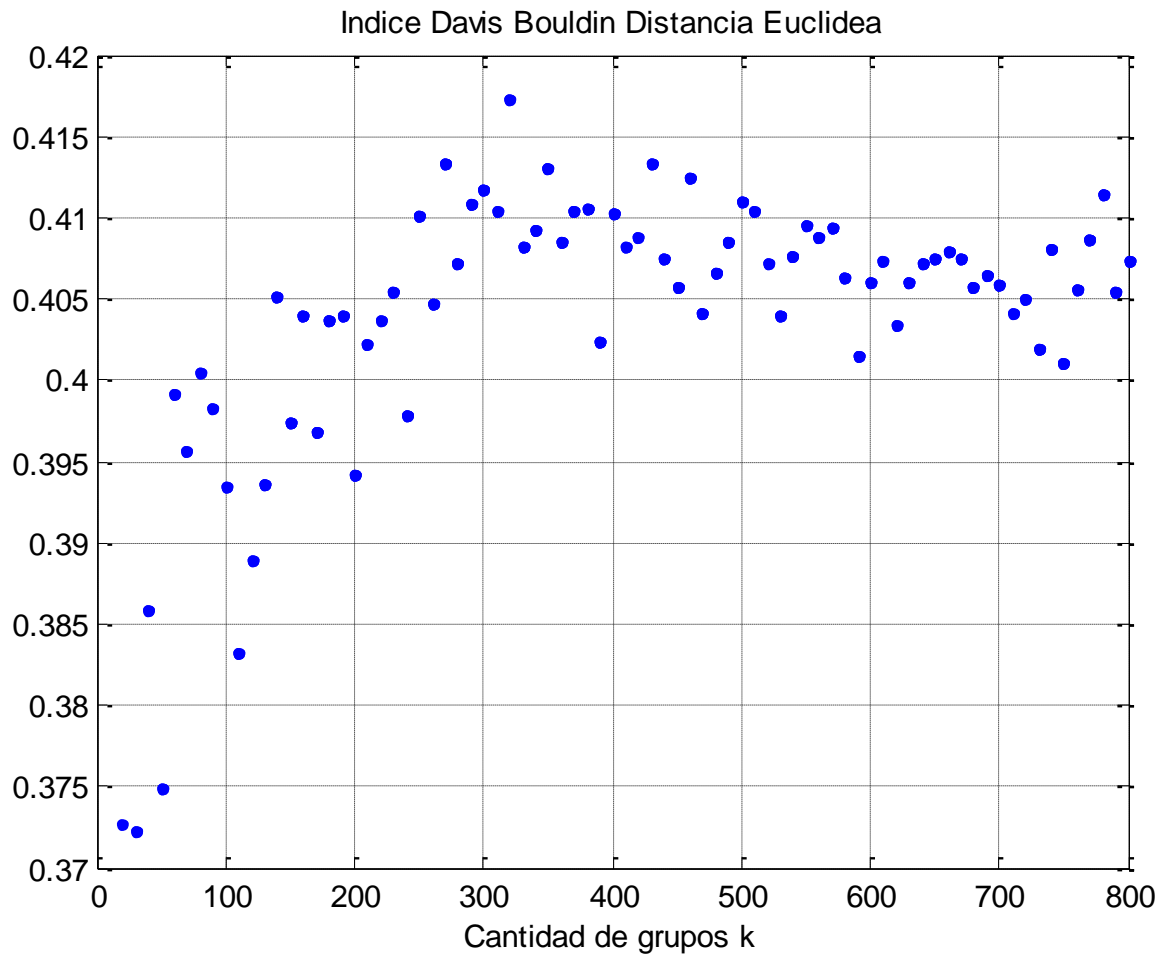
4.3.2.2 Selección del número k de grupos adecuados

Para la selección adecuada del número de centroides, se selecciona la norma Euclídea para calcular la distancia, y se realiza el proceso de agrupación variando la cantidad de grupos a crear. A partir de las estrategias para evaluar la metodología de agrupación, mostrada en la sección 0, se selecciona la mejor cantidad de prototipos que representan el sistema.

a. Evaluación del índice Davis Bouldin

Para variar la cantidad de prototipos que representan el sistema, desde 20 hasta 800, se encuentra que el índice Davis Bouldin para cada agrupamiento, presenta el comportamiento mostrado en la Figura 4.3

Figura 4.3: Comportamiento del índice Davis Bouldin para cada uno de los agrupamientos

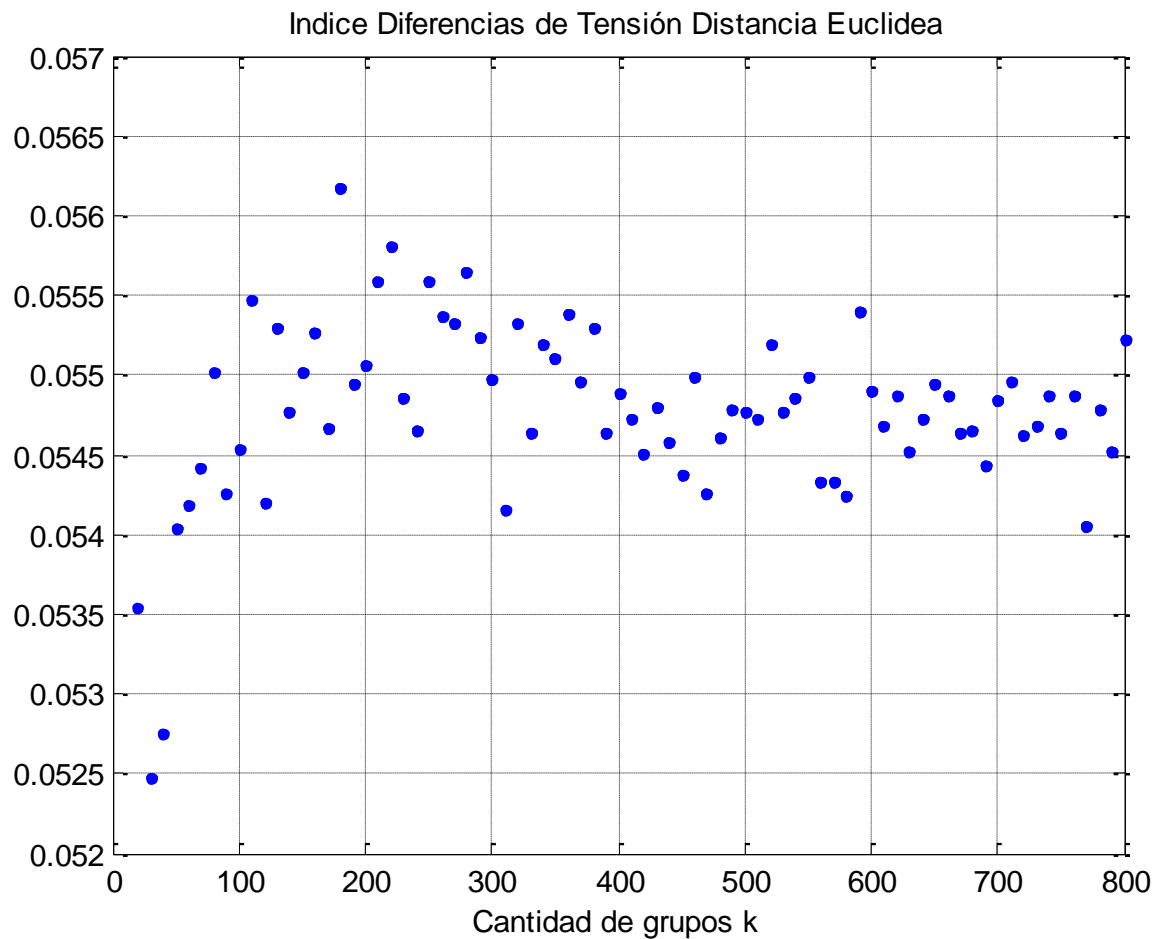


Al analizar el comportamiento de este índice, se observa que el error se estabiliza después de seleccionar 400 grupos.

b. Evaluación del índice de diferencias de tensión

Si se varía la cantidad de prototipos a crear para representar el sistema y con el mismo rango de variación de k utilizado en el índice Davis Bouldin, entonces el comportamiento del índice de diferencias de tensión obtenido se ilustra en la Figura 4.4

Figura 4.4: Comportamiento del índice de diferencias de tensión para cada uno de los agrupamientos

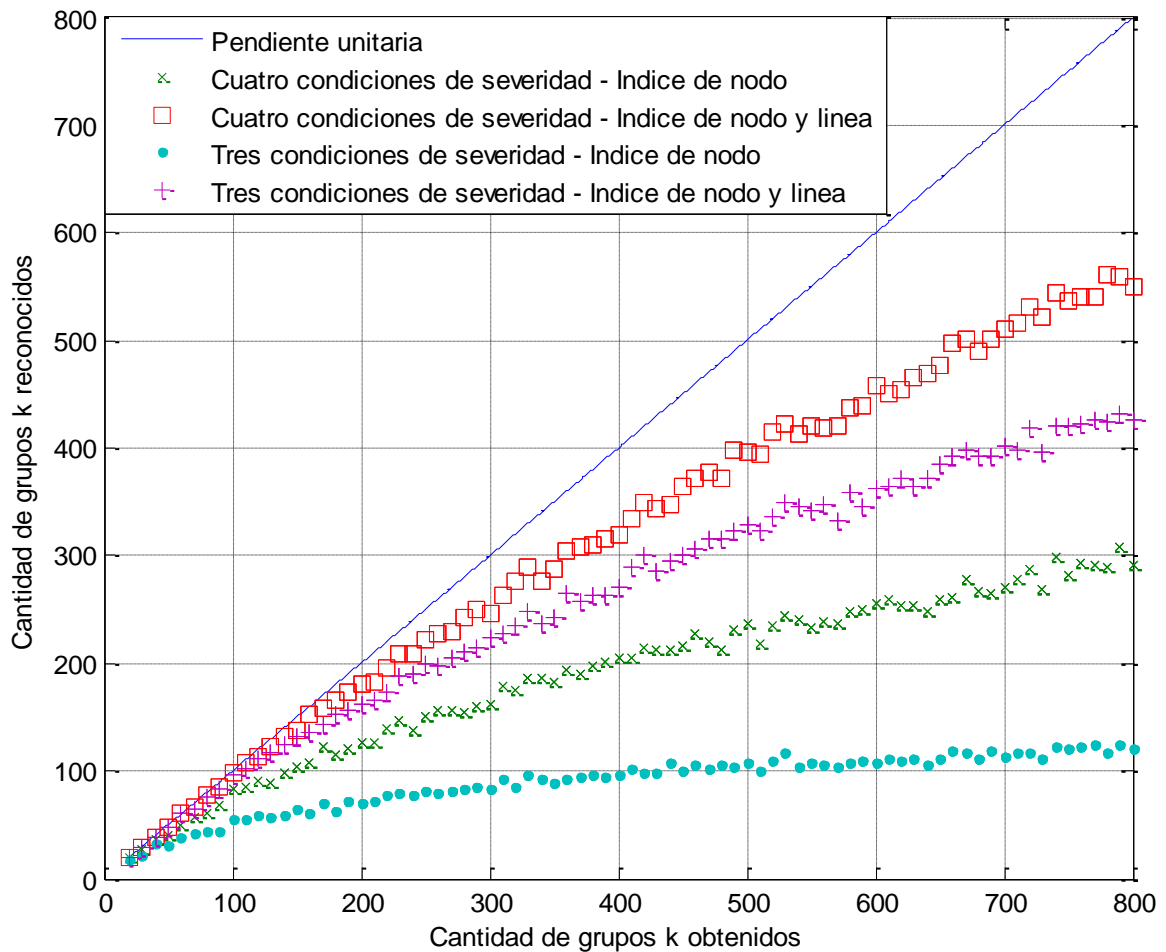


Por lo tanto, con el comportamiento de este índice se observa que no presenta grandes variaciones, cuando se seleccionan más de 400 grupos.

c. Evaluación mediante la definición de rangos cualitativos de los índices de estabilidad de tensión

Con este método de evaluación, se consideran para cada una de las condiciones de operación, diferentes zonas de severidad, tres o cuatro rangos de criticidad para los índices de estabilidad. El comportamiento mostrado en la Figura 4.5, ilustra la cantidad de prototipos representativos cuando se consideran diferentes escenarios así:

- Considerando tres rangos de criticidad para los índices de barra.
- Considerando tres rangos de criticidad para los índice de barra y los índices de línea.
- Considerando cuatro rangos de criticidad para los índices de barra.
- Considerando cuatro rangos de criticidad para los índices de barra y los índices de línea.

Figura 4.5: Comportamiento cualitativo según las zonas de severidad definida para cada uno de los agrupamientos

Los índices utilizados para definir las zonas de severidad son el índice SVSI para el comportamiento de las barras y el índice LQP para el comportamiento de los índices de línea. Cuando se definen 4 intervalos, los rangos de criticidad definidos para cada uno de los índices se muestran en la Tabla 4.1

Tabla 4.1: Rangos de criticidad para definir las zonas de severidad

Índice de barra SVSI		Índice de línea LQP	
Rango de criticidad	Descripción	Rango de criticidad	Descripción
$SVSI \leq 0.5$	Seguro	$LQP \leq 0.5$	Seguro
$0.5 < SVSI \leq 0.75$	Alerta	$0.5 < LQP \leq 0.75$	Alerta
$0.75 < SVSI \leq 0.9$	Critico	$0.75 < LQP \leq 0.9$	Critico
$SVSI > 0.9$	Muy critico	$LQP > 0.9$	Muy critico

Considerando el índice SVSI, como un buen indicador de la estabilidad para este sistema, se tiene que el número significativo de prototipos no varía significativamente después de 350 grupos.

4.3.3 Determinación del mejor agrupamiento

A partir de los tres criterios de evaluación, el índice Davis Bouldin, el índice de diferencias de tensión y las clases representativas cualitativamente al definir las zonas de severidad, se determina que el índice Davis Bouldin se estabiliza a partir de la creación de 400 grupos. Una situación similar se presenta para el error de diferencias de tensión. Finalmente, al definir rangos de criticidad se observa que no se obtiene mayor número de grupos significativos al aumentar la cantidad de grupos de 350. De acuerdo a lo anterior, para el sistema IEEE de 14 barras, el mejor agrupamiento obtenido con el algoritmo *k means*, se obtiene al crear 400 grupos y utilizar la norma Euclídea, para calcular la distancia, como parámetros de entrada al algoritmo.

4.4 Reconocimiento de áreas críticas de estabilidad

Con la agrupación realizada con la técnica *k means*, se tiene una base de datos con diferentes condiciones de operación representadas por diferentes prototipos. Por lo tanto, se puede utilizar esta base de datos para desarrollar una herramienta de clasificación de nuevas condiciones de operación.

La técnica de clasificación *knn*, aplicada al sistema IEEE de 14 barras para reconocimiento de áreas críticas de estabilidad, presenta los siguientes resultados.

4.4.1 Conjunto de entrada

La base de datos utilizada como conjunto de entrada, para el entrenamiento del algoritmo *knn*, es la misma utilizada para el proceso de agrupación. Esta base está conformada por aquellas condiciones de operación en donde al menos uno de los índices de línea LQP o índice de barra SVSI, se encuentre catalogado como crítico como resultado de aplicar la definición de zonas de severidad.

Debido a que se requiere un proceso robusto, que funcione con las variables suficientes del sistema de potencia, y a la necesidad de un reconocimiento en línea de la estabilidad en tensión, como descriptores de entrada solo se seleccionan las tensiones en las barras en magnitud y ángulo y los valores de potencia activa y reactiva generada.

Adicionalmente, para el proceso de evaluación y parametrización del algoritmo, se agrega una columna que contiene la información del grupo asignado a esta condición de operación, mediante la técnica *k means*, tal y como se muestra en la **Tabla 4.2**

Tabla 4.2: Descriptores utilizados para describir las condiciones de operación en el proceso de clasificación

Condición de operación	Tensiones en las barras		Potencia en los generadores		Clase
	Magnitud	Angulo	Pgen	Qgen	
X_k	(1-m)	(1-m)	(1-g)	(1-g)	k

Donde m es la cantidad de barras del sistema, g la cantidad de generadores conectados al sistema y k la clase a la cual pertenece la condición de operación obtenida por el proceso de agrupación *k means*.

4.4.2 Selección de los parámetros de la técnica de clasificación

Para la determinación de los parámetros de la técnica *knn*, se realiza una prueba extensiva variando la cantidad de condiciones de operación que conforman el vecindario de decisión y la forma de ponderar la distancia a cada uno de las variables.

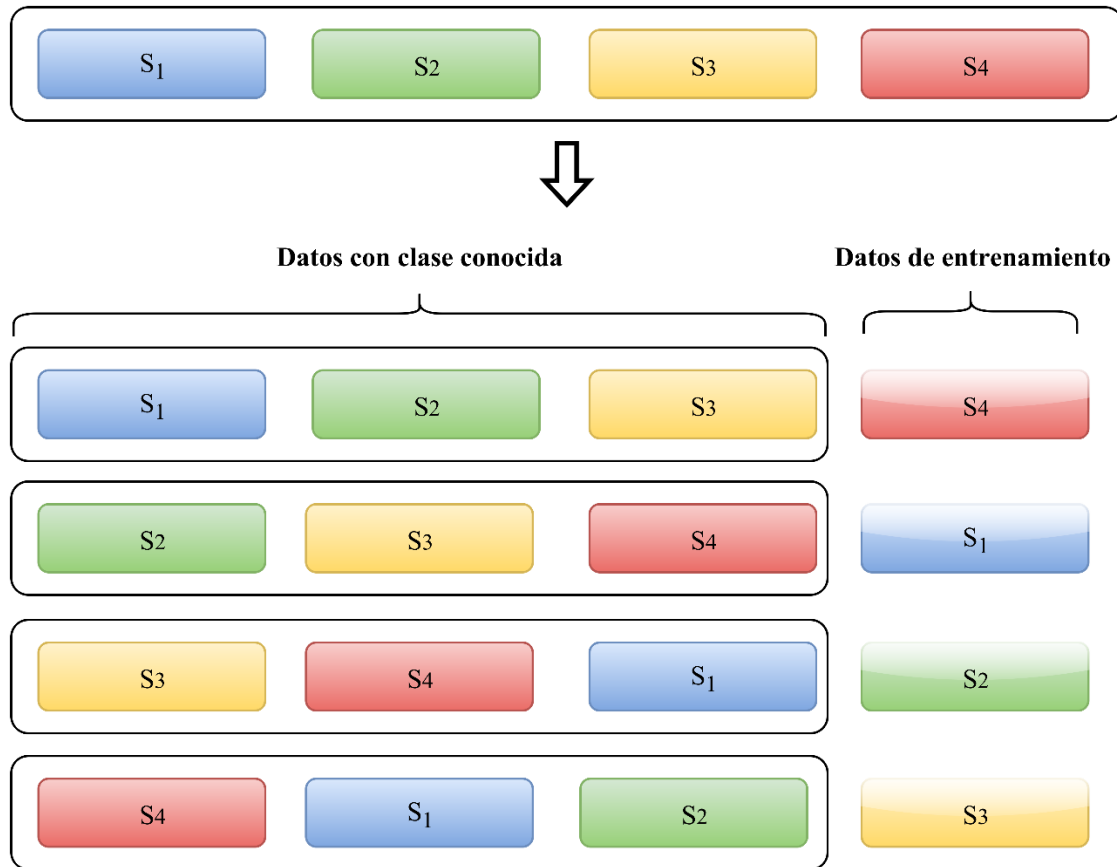
La forma de evaluación de cada una de las variaciones, se realiza utilizando la técnica de validación cruzada y midiendo el desempeño mediante las expresiones (3.17) y (3.18).

4.4.3 Validación cruzada

Una forma de evaluar el desempeño de la técnica de clasificación, es mediante la técnica de validación cruzada. Para este caso, de las 32265 condiciones de operación que conforman la base de datos, se crean cuatro subconjuntos. Cada uno de los subconjuntos contienen igual cantidad de condiciones de operación y la pertenencia de una condición de operación a cada subconjunto es tomada de forma aleatoria. Adicionalmente, cada una de las condiciones de operación está caracterizada por los descriptores mostrados en la Tabla 4.2.

Así, cada uno de los subconjuntos se utiliza como datos de entrada al clasificador, y el resto de datos servirá como base de datos con condiciones de operación identificadas previamente. En la Figura 4.6 se muestra la división de la base de datos original, y la combinación realizada para el entrenamiento del clasificador.

Figura 4.6: Subconjuntos creados para la evaluación del clasificador mediante validación cruzada



4.4.4 Modificaciones al algoritmo básico

Las modificaciones realizadas al algoritmo básico de clasificación, se prueban con cada uno de los subconjuntos creados mediante la técnica de validación cruzada. Al variar la cantidad k de vecinos que conforman el vecindario, se evalúa el desempeño de cada modificación para así determinar su efectividad.

Por consiguiente, se consideran los siguientes escenarios:

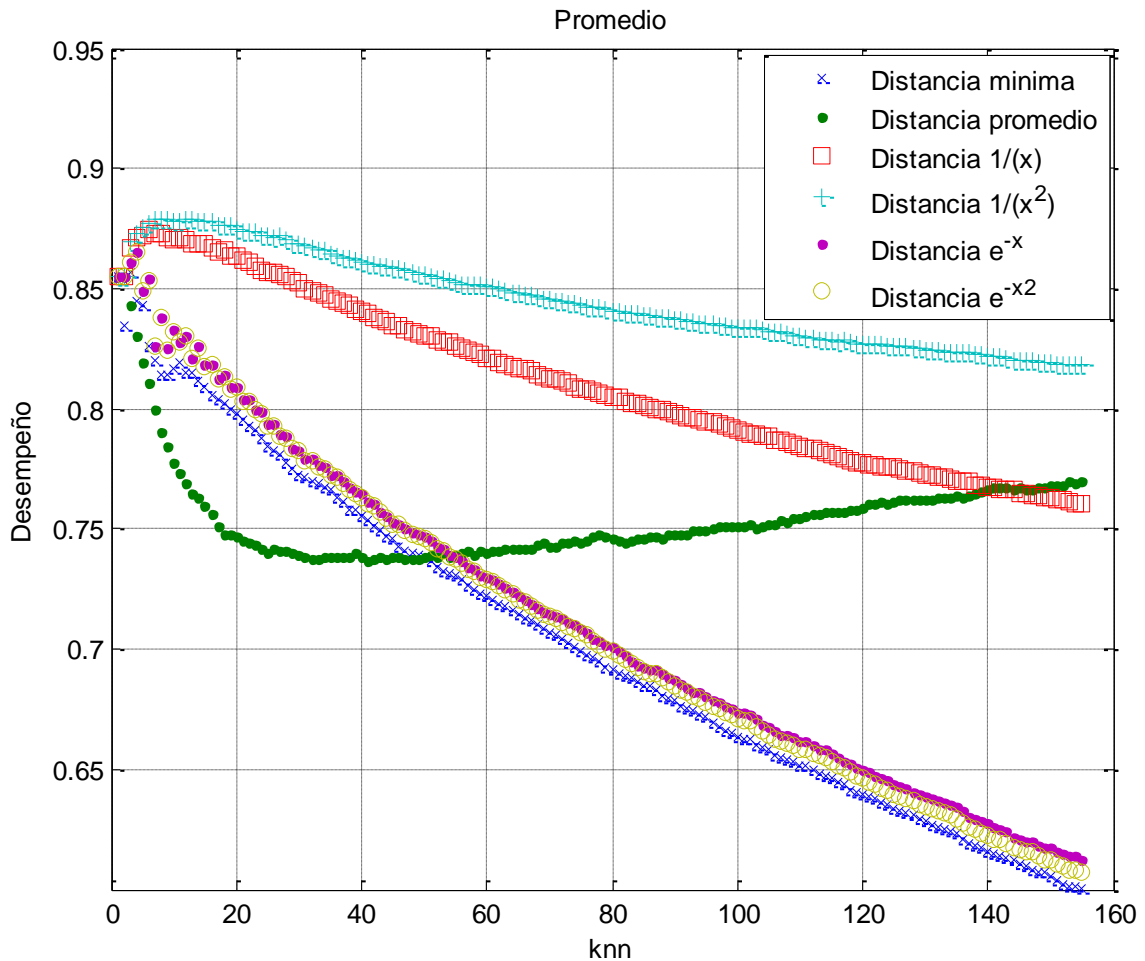
- Algoritmo básico *knn*.
- Algoritmo *knn* con modificación de distancia promedio.
- Algoritmo *knn* con ponderación de distancia utilizando la función inversa $1/x$
- Algoritmo *knn* con ponderación de distancia utilizando la función inversa $1/x^2$

e. Algoritmo *knn* con ponderación de distancia utilizando la función inversa $1/e^x$

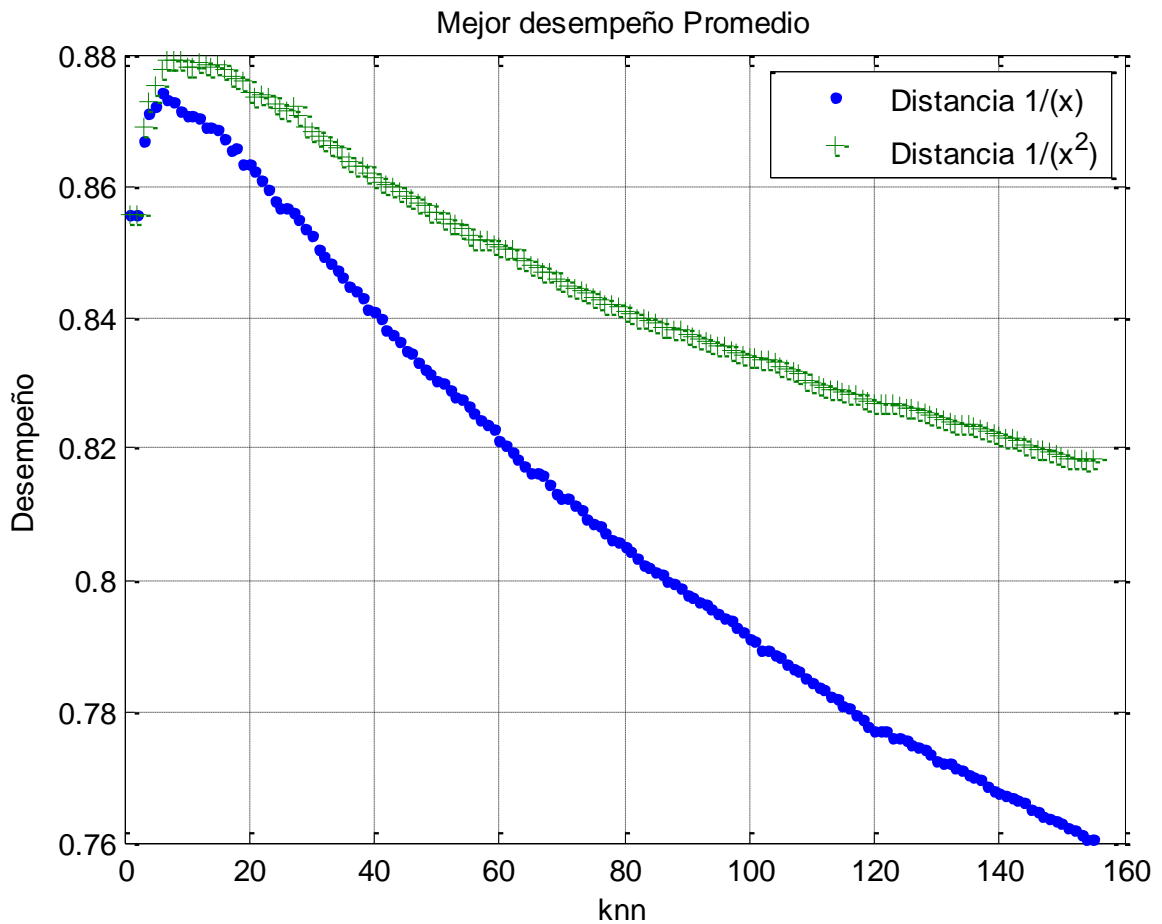
f. Algoritmo *knn* con ponderación de distancia utilizando la función inversa $1/e^{x^2}$

En la Figura 4.7 se muestra el comportamiento del desempeño cuando se realizan las modificaciones al algoritmo, se varía la cantidad de vecinos que pertenecen al vecindario, y se utiliza cada uno de los subconjuntos; este desempeño se evalúa con la expresión (3.18).

Figura 4.7: Desempeño del clasificador



El desempeño promedio del clasificador, muestra que a medida que la cantidad de condiciones de operación que conforman el vecindario aumenta, el desempeño desmejora para todas las condiciones, excepto en la condición de distancia promedio, que se comporta de forma aleatoria, y que después de cierto número k de vecinos, tiende a estabilizarse. Sin embargo, se observa que las funciones con mejor comportamiento, son las funciones inversas ponderadoras de distancia $1/x$ y $1/x^2$, tal y como se observa en la Figura 4.8.

Figura 4.8: Desempeño del clasificador con los dos mejores parametros

4.4.5 Parametrización del algoritmo k vecinos más cercanos

Teniendo en cuenta el desempeño mediante la técnica de validación cruzada, se observa un desempeño superior al 78% cuando se utiliza la función $1/x$, y un desempeño superior al 82% cuando se utiliza la función $1/x^2$. De acuerdo a lo anterior, para el sistema IEEE de 14 barras, el mejor desempeño de la metodología de clasificación mediante el algoritmo knn , se obtiene al parametrizar con una cantidad k de vecinos igual a ocho y una función de ponderación de distancia con comportamiento $1/x^2$.

4.4.6 Evaluación del clasificador *knn* parametrizado

Con los parámetros ingresados, se prueba la efectividad del clasificador con un 14.82% del total de la base de datos, correspondiente a 4840 datos. Estos datos son escogidos completamente de forma aleatoria.

A partir de los resultados, son etiquetadas correctamente 4298 condiciones de operación, por lo tanto, el clasificador presenta un error del 11.12%.

Uno de los factores determinantes de la metodología propuesta, es el tiempo que tarda la técnica en clasificar la condición de operación actual; A partir de los resultados, el tiempo promedio para la clasificación de cada uno de los 4840 datos es de 9.2 ms con una desviación estándar de 0.29776 ms. Estos resultados fueron obtenidos en un equipo de cómputo con las características mostradas en la Tabla 4.3.

Tabla 4.3. Especificaciones del equipo de computo.

Procesador	Intel® Core™ i5-3470 CPU @ 3.20 GHz
Sistema Operativo	Windows 7 Professional
Fabricante	Hewlett-Packard Company
Modelo	HP Compaq Pro 6300 MT
Memoria Instalada (RAM)	8,00 Gb
Tipo de sistema	Sistema Operativo de 64 bits

4.4.7 Comparación de la técnica de clasificación *knn* con la máquina de soporte vectorial

Para la clasificación de una condición de operación se pueden emplear diferentes técnicas. La técnica de clasificación, basada en el algoritmo *knn* se estudia en esta tesis, sin embargo, otras alternativas son empleadas como redes neuronales, árboles de decisión, mezclas finitas, máquinas de soporte vectorial, entre otras.

Una de las recomendaciones para trabajos futuros realizadas en (Pérez Londoño, 2013), es la evaluación de una técnica de clasificación alternativa para una condición de operación del sistema eléctrico de potencia.

Con base en lo anterior, se tienen ciertas características importantes para realizar comparaciones entre la técnica de clasificación basada en el algoritmo *knn*, propuesta en esta tesis y la técnica de clasificación mediante máquinas de soporte vectorial, propuesto en la tesis previa. Estas características son:

- a. La técnica utilizada en esta tesis se basa en el algoritmo *knn*. La técnica de clasificación utilizada en la tesis previa, se basa en la máquina de soporte vectorial; ésta se parametriza mediante la técnica del algoritmo genético de Chu Beasley. Esta modificación se realiza para probar una técnica alternativa de clasificación.
- b. La matriz de variación de carga para generar la base de datos es diferente. Por lo tanto, la base de datos con situaciones críticas, con base en la definición de zonas de severidad, utilizada para el entrenamiento y la efectividad del algoritmo también es diferente. En esta tesis se utiliza una matriz con 32265 condiciones de operación, mientras que en la tesis previa se utiliza una matriz con 49833 condiciones de operación
- c. Los descriptores utilizados para la clasificación en esta tesis son: La magnitud y ángulo de tensión para cada una de las barras y la potencia activa y reactiva generada por cada uno de los generadores. Los descriptores utilizados para la técnica de clasificación en la tesis previa y que dieron mejor resultado son: La magnitud y ángulo de tensión para cada una de las barras y los flujos de potencia por las líneas.
- d. La metodología se ejecuta en un sistema de cómputo con características diferentes.
- e. La prueba de la técnica de clasificación utilizada en esta tesis contiene un 14.82% de la base de datos original, es decir, 4840 condiciones de operación. La prueba de la técnica de clasificación utilizada en la tesis previa contiene un 20.06% de la base de datos original, es decir, 10000 condiciones de operación.
- f. La selección de la cantidad de prototipos representativos en esta tesis es de 400 grupos. La selección de la cantidad de prototipos representativos en la tesis previa es de 250 grupos.

A partir de los resultados, se observa que la técnica de clasificación utilizando máquina de soporte vectorial, para el sistema de prueba IEEE de 14 barras, presenta una precisión en el mejor de los casos de 99.60%. Para la prueba con la técnica de clasificación utilizando el algoritmo *knn*, utilizada en esta tesis, es del 88.88%.

5. Conclusiones y recomendaciones

5.1 Conclusiones

Debido a la importancia de la estabilidad en los sistemas eléctricos de potencia, existe la necesidad de identificar las áreas del sistema con problemas de estabilidad en tensión. Luego de identificar las áreas con problemas, se debe establecer protocolos para ejercer las maniobras adecuadas de control sobre el sistema eléctrico de potencia. En consecuencia, se observa que utilizando técnicas de minería de datos y aprendizaje de máquina, se pueden caracterizar las diferentes zonas del sistema de potencia, y además, obtener prototipos que representan el comportamiento del sistema de potencia, ante diversas condiciones de operación.

Una de las formas identificadas para describir el sistema eléctrico de potencia, es a través de los índices de estabilidad de los elementos que lo componen y las variables de estado. Además, con las nuevas herramientas de medición al sistema, como las PMU, se pueden obtener mediciones en tiempo real; con estas mediciones se puede clasificar el comportamiento del sistema y en base a los protocolos previamente identificados, prevenir o corregir el problema de estabilidad de tensión en un área del sistema eléctrico.

Con la metodología propuesta se sugiere crear prototipos que representan al sistema de potencia bajo diferentes condiciones de carga y/o contingencia. De esta forma, se definen las zonas sensibles en cuanto a la estabilidad de tensión.

Las diversas formas de operación del sistema se obtienen mediante la técnica de agrupación *k means*, una técnica ampliamente utilizada y de fácil implementación. La técnica de agrupación requiere una metodología adecuada de evaluación, para obtener parámetros óptimos y asegurar su eficiencia. Los métodos de evaluación de la metodología de agrupación, aparte de índices matemáticos, requiere un método adaptado al problema físico; esta es la razón para utilizar métodos de evaluación que tienen en cuenta el sentido real del planteamiento, como son los índices de diferencias de tensión y la evaluación cualitativa basada en zonas de severidad.

Se observa adicionalmente, que la base de datos de entrenamiento y los parámetros iniciales, con los que se ejecuta la técnica de agrupamiento *k means*, tienen mucha influencia en la forma como se determinan los prototipos que describen el sistema. Por lo descrito anteriormente, si se cuenta con una base de datos, la cual presenta variables concentradas alrededor de un punto, se observa que el algoritmo presenta un menor desempeño al asignar las condiciones de operación que pertenecen a cada uno de los prototipos, mientras que si la base de datos presenta condiciones que se condensan en diferentes puntos, se observa que el desempeño mejora considerablemente.

La necesidad de agrupar diferentes condiciones operación, surge debido a la gran cantidad de condiciones que se pueden presentar en la operación, por lo tanto, si se simplifican los protocolos de control, que puede ejercer el operador de red sobre los elementos que componen el sistema de potencia, se pueden evitar daños masivos sobre el mismo, incluso un posible colapso de tensión.

El reconocimiento en línea de condiciones de operación, es esencial para establecer adecuadamente los criterios de control a ejercer sobre el sistema de potencia. De acuerdo a esto, la metodología de clasificación de condiciones de operación mediante la técnica *knn* propuesta en esta tesis, presenta una fácil implementación en el sistema con un desempeño alto. Al utilizar un método que presenta bajo costo computacional, es posible identificar ágil y eficientemente las nuevas condiciones de operación del sistema de potencia. Esto facilita a definición de maniobras preventivas o de control necesarias para evitar un daño en los equipos o un colapso de tensión.

La forma correcta de parametrizar el algoritmo de clasificación *knn*, se evalúa mediante técnicas de alta validez estadística, con el fin de obtener un gran desempeño al clasificar, y que la técnica tenga validez. A partir de los resultados obtenidos, la técnica de clasificación presenta altos índices de desempeño para el problema planteado sobre el sistema de prueba IEEE de 14 barras.

Al comparar la técnica de clasificación basada en el algoritmo *knn* propuesta en esta tesis, y la técnica de clasificación basada en máquinas de soporte vectorial, se puede concluir que en la aplicación al sistema de prueba tipo IEEE de 14 barras, el *knn* presenta mayor confusión al clasificar las condición de operación. Sin embargo, la técnica de clasificación propuesta en esta tesis presenta un costo computacional menor en comparación con la máquina de soporte vectorial.

5.2 Recomendaciones

Como recomendaciones para aplicar la metodología de reconocimiento de áreas críticas con problemas de estabilidad de tensión, se tienen principalmente consideraciones al momento de ejecutar los algoritmos propuestos y la creación de la base de datos a utilizar.

La base de datos debe contener diversas condiciones de operación, en donde el sentido físico del problema no se pierda. Se debe considerar rangos realistas de variación de carga de los nodos.

Para una correcta aplicación de la metodología propuesta, se deben identificar adecuadamente los descriptores o variables de cada condición de operación que harán parte en el proceso de agrupación, debido a las variaciones erróneas en la respuesta por inclusión de variables que no aportan información necesaria del problema.

Debido a la necesidad de la rápida respuesta del clasificador de condiciones de operación, para este caso, es muy importante reconocer cuales son los descriptores suficientes y necesarios para implementar la técnica. Existen variables adicionales que no aportan demasiada información, consumen tiempo de ejecución.

Para la parametrización de las técnicas de agrupación y clasificación, se deben tener métodos adecuados de evaluación, en los cuales, no solo se involucren métodos matemáticos, sino se involucre el sentido físico del problema.

5.3 Trabajos futuros

Como recomendaciones para trabajos futuros en el tema se tienen:

Aplicación y evaluación de la metodología en sistemas de prueba con mayor complejidad; evaluar resultados, comparar eficiencias y ajustar correctamente las técnica de agrupación y clasificación; con base en los resultados, proponer la implementación de la metodología en sistemas reales.

Evaluación de la parametrización de la técnica de agrupación basada en el algoritmo *knn*, como la combinación entre la metodología de ponderación de distancia y la metodología de ponderación de variables.

Bibliografía

- Amraee, T., Ranjbar, A. M., Mozafari, B., & Sadati, N. (2007). An enhanced under-voltage-shedding scheme to provide voltage stability. *Electric Power Systems Research*, 1038-1046.
- Balamourgan, V., Sidhu, T., & Sachdev, M. (2004). Technique for online prediction of voltage collapse. *IEE Proc-Generation, Transmission, Distribution*, 151, 453-460.
- Congreso de la Republica de Colombia. (1994). *Ley de servicios públicos domiciliarios: Ley 142 de 1994*. Bogotá: Señal Editora.
- Congreso de la República de Colombia. (1994). *Ley Eléctrica: Ley 143 de 1994*. Bogotá: Señal Editora.
- Davies, D., & Bouldin, D. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 95-104.
- Esqueda Elizondo, J. J., & Palafox Maestre, L. E. (2005). *Fundamentos para el procesamiento de imágenes*. Mexicali, Baja California: UABC.
- Fix, E., & Hodges, J. (1970). *Discriminatory analysis, nonparametric discrimination*. Randolph Field.
- Gutierrez Gallego, J., Pérez Londoño, S., & Mora Flórez, J. J. (2010). Efficient adjust of a learning based fault locator for a power distribution systems. *Transmission and Distribution Conference and Exposition: LatinAmerica (T&D-LA)*, 774-792.
- Helton, J., & Davis, F. (2003). Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems. *Reliability Engineering and System Safety*, 23-69.
- Kessel, P., & Glavitsch, H. (1986). Estimating the Voltage Stability of a Power System. *IEEE Transactions on Power Delivery*.
- Kundur, P. (1994). *Power System Stability and Control*. New York, USA: McGraw-Hills, Inc.
- Machowski, J., Bialek, J., & Bumby, J. (2008). *Power Systems Dynamics: Stability and Control*. United Kingdom: Wiley 2nd Edition.
- MacQueen, J. (1968). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.

- Mena Kodsi, S. K., & Cañizares, C. A. (2003). *Modeling and Simulating of IEEE 14 bus system with facts controllers*. Waterloo.
- Milano, F. (2001). *Power System Analysis Toolbox – PSAT*. Dublin.
- Moghavvemi, M., & Faruque, M. (1999). Power system security and voltage collapse. a line outage based indicator for prediction. *Electric Power and Energy System* , 455-461.
- Moghavvemi, M., & Faruque, O. (1998). Real-time contingency evaluation and ranking technique. *IEE Proceedings Generation, Transmission and Distribution*.
- Moghavvemi, M., & Faruque, O. (2001). Technique for assessment of voltage stability in ill-conditioned radial distribution network. *IEEE Power Engineering Review*, 58-60.
- Mora Floréz, J. J. (2006). *Localización de faltas en sistemas de distribución de energía eléctrica usando métodos basados en el modelo y métodos basados en el conocimiento*. Tesis doctoral, Girona, España.
- Moreno, F. (2004). *Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano*. Alicante: Departamento de lenguajes y sistemas informáticos.
- Moujahid, A., & Inza I, L. P. (2008). Clasificadores knn. *Departamento de Ciencias de la Computación e Inteligencia Artificial*.
- Nolasco Contreras, E. M. (2012). *Estimación de estados de sistemas eléctricos de potencia considerando mediciones fasoriales*. Mexico DF.
- NYSERDA. (2010). *Real-time applications of phasor measurement units PMU for visualization, reactive power monitoring and voltage stability protection*. Reporte, New York.
- Pérez Londoño, S. M. (2013). *Propuesta para la identificación y monitoreo de áreas débiles con relación a la estabilidad de tensión en sistemas eléctricos de potencia*. Manizales.
- Pérez Londoño, S. M., Rodríguez García, L. F., & Olivar, G. (2014). A simplified voltage stability index (svsi). *International Journal of Electrical Power and Energy Systems*, 806-813.
- Russell, S., & Norvig, P. (2003). Artificial intelligence: a modern approach. *Pearson Education*.
- Tou JT, G. (1997). *Pattern recognition principles*. Addison-Wesley Publication Company Reading.

- U S -Canada Power System Outage Task. (2011). *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. General Books.
- Wang, B., Zeng, Y., & Yang, Y. (2008). *Generalized nearest neighbor rule for pattern classification*. China.
- Yesuratnam, G., & Thukaram, D. (2007). Congestion management in open Access based on relative electrical distances using voltage stability criteria. *Electric Power Systems Research*, 1608-1618.